Electronics Below 10 nm

Konstantin Likharev
State University of New York, Stony Brook, NY 11794-3800

Abstract

This chapter reviews prospects for the development and practical introduction of ultrasmall electron devices, including nanoscale field-effect transistors (FETs) and single-electron transistors (SETs), as well as new concepts for nanometer-scalable memory cells. Physics allows silicon FETs to be scaled down to ∼3 nm gate length, but below ∼10 nm the devices are extremely sensitive to minute (sub-nanometer) fabrication spreads. This sensitivity may send the fabrication facilities costs (high even now) skyrocketing, and lead to the end of the Moore Law some time during the next decade. Lithographically defined SETs can hardly be a panacea, since the critical dimension of such transistor (its single-electron island size) for the room temperature operation should be below ∼1 nm. Apparently, the only breakthrough that would allow to make 1-nm-scale electron devices practical, would be the introduction of "CMOL" hybrid integrated circuits that would feature, in addition to an advanced CMOS subsystem, a layer of ultradense molecular electron devices. These devices would be fabricated by chemically-assisted self-assembly from solution on few-nm-pitch nanowire arrays connecting them to the CMOS stack. Due to the finite yield of molecular devices and their sensitivity to random charged impurities, this approach will require a substantial revision of integrated circuit architectures, ranging from defect-tolerant versions of memory matrices and number crunching processors to more radical solutions like hardware-implemented neuromorphic networks capable of advanced image recognition and more intelligent information processing tasks.

Key words: nanoelectronics, electron devices, memory cells, logic

1. Introduction

The phenomenal success of semiconductor electronics during the past three decades was based on scaling down of silicon field-effect transistors (MOSFETs) and the resulting increase of density of logic and memory chips. The most authoritative industrial forecast, the International Technology Roadmap for Semiconductors [1] predicts that this exponential ("Moore-Law") progress of silicon MOSFETs and integrated circuits will continue at least for the next 15 years. By the end of this period, devices with 10-nm minimum features (transistor gate length L) should become commercially available.

The prospects to continue the Moore Law beyond the 10 nm frontier are more uncertain, and it is very important to understand them. This chapter is an attempt at a review of the recent research results related to the scaling prospects of silicon MOSFETs and possible alternatives to this technology. I will begin the review with a discussion (Sec. 2) of advanced field effect transistors, with a focus on their most prospective variety: double-

gate SOI MOSFETs. In Sec. 3, single electron transistors and other Coulomb-blockade-based devices will be discussed. (A brief review of single- and few-electron memory cells and some other new memory ideas is also included in that section.) This will naturally lead us to a discussion of molecular devices and prospects of their self-assembly and hybridization with CMOS technology (Sec. 4). Finally, in Conclusion (Sec. 5) I will try to summarize the basic problems facing future integrated circuits beyond the 10 nm frontier, and possible ways of their solution.

2. Advanced Field Effect Transistors

2.1. *Bulk and SOI MOSFETs*

Bulk silicon MOSFETs (for their detailed description, see, e.g., Refs. 2-5) are extremely versatile electron devices combining a (relatively) easy fabrication with very high performance in a broad variety of logic and memory circuits. Moreover, the devices are scalable to deep-submicron range. This powerful combination has allowed the bulk MOSFET devices to serve as the work horse of the leading electronic technology, CMOS, for more than 30 years. However, as the bulk MOSFETs enter the sub-100-nm range, their further scaling runs into several problems, including short-channel effects and gate oxide leakage – see Ref. 6 for an extensive review of these issues. Despite the recent experimental demonstrations of several bulk transistors with gate length below 20 nm [7-14], performance of these prototypes is far from perfect.

There is a growing consensus (see, e.g., Chapter 1 of this collection) that reaching high performance (good saturation at high ON current and high ON/OFF ratio) below 20 nm will require the use of advanced FETs, primarily double-gate MOSFETs with thin, undoped silicon-on-insulator (SOI) channel connecting highly doped source and drain. The main reasons in favor of this choice are as follows:

1 Such device is a close approximation to what may be called the "ultimate MOSFET", because two gates allow a very effective control of the electrostatic potential of the channel, and hence the carrier transport. (Similar devices with single-gate [15, 16] loose to double-gate devices in scalability, though are certainly preferable to bulk MOSFETs.)

2 Although the fabrication of double-gate transistors is certainly more complex than that of the usual bulk MOSFETs, they have already been implemented in various geometrical versions, including planar [17-24], fin-type [25-33] and vertical [34-37] geometries.

Because of these reasons, the double-gate MOSFETs have become a focus of recent theoretical efforts to understand MOSFET scaling laws and limits [38-61]. These analyses are based on a variety of models and calculation techniques, but give results in the same ballpark. I will present the most recent results [59, 61] based on a model (Fig. 1) that I believe provides the best trade-off of analysis simplicity and result accuracy.

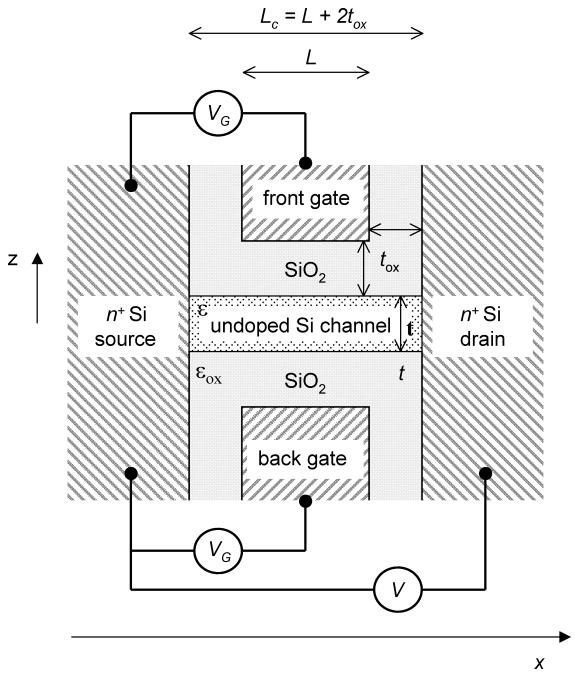

Fig. 1. A simple model of double-gate MOSFETs with ultrathin intrinsic channel. Notice the difference between the gate length L and channel length L_c.

2.2. *Model*

In this model of a nanoscale n-MOSFET, the channel thickness$t$ is assumed to be so small that the lateral electron confinement (in direction z perpendicular to the channel plane) considerably increases the effective potential energy

$$U_{ef}(x,y) = E_z - e\langle\Phi\rangle(x,y) \tag{1}$$

of 2D motion of electrons in the channel ($e > 0$). Here

$$\langle\Phi\rangle = \int_{-t/2}^{+t/2} \Phi(x,y,z)\,|\psi(x,y,z)|^2 dz \tag{2}$$

is the effective 2D value of the 3D electrostatic potential Φ. Calculations [86] show that if $t \lesssim 3$ nm, while the channel length L_c and width W are much larger than t, the 3D electron wavefunction ψ may be factored as

$$\psi \approx \frac{1}{\sqrt{2t}} \cos\frac{\pi z}{2t} \sum_{E_x,k_y} \psi(x) e^{ik_y y}, \tag{3}$$

so that the confinement energy

$$E_z = \frac{\pi^2\hbar^2}{2m_h t^2}, \tag{4}$$

where m_h (close to m_0 for silicon) is the heavy electron mass. For the value $t = 2$ nm, accepted in most illustrations below, E_z is quite considerable (~ 0.1 eV). (The channel is assumed to be in the [102] direction of the silicon crystal, so that the only two valleys with heavy mass in z direction participate in transport; for four other valleys with light mass $m_l \approx 0.19\ m_0$ in z direction, the confinement energy is $m_h/m_l \approx 5$ times higher and for reasonable values of applied voltages these valleys are not populated.)

The functions $\psi(x)$ and $\Phi(x,y)$ (if $W \gg L_c$, the y-dependence of Φ is negligible) may be found by a solution of, respectively, the 1D Schrödinger equation

$$-\frac{\hbar^2}{2m_l}\frac{\partial^2\psi(x)}{\partial x^2} + U(x)\psi(x) = E_x\psi(x), \tag{5a}$$

$$U(x) \equiv U_{eq}(x,y) - \frac{\hbar^2 k_y^2}{2m_l}, \tag{5b}$$

and the 2D Poisson equation

$$\frac{\partial^2\Phi}{\partial x^2} + \frac{\partial^2\Phi}{\partial y^2} = \begin{cases} en(x,z)/\varepsilon\varepsilon_0, & |z| \le t/2, \\ 0, & t/2 \le |z| \le t/2 + t_{ox}. \end{cases} \tag{6}$$

Since $n(x,z)$ is the full 3D density of electrons in the channel:

$$n(x,z) = |\psi(x,y,z)|^2 = \frac{1}{2t}\cos^2\frac{\pi z}{2t} \sum_{E_x,k_y} |\psi(x)|^2, \tag{7}$$

equations (5ab) and (6) are generally coupled and should be solved self-consistently. After that, drain current density (per unit channel width) may be calculated as

$$J = 4\frac{e\hbar}{m_l}\mathrm{Im} \sum_{E_x,k_y} \psi^*(x)\frac{d}{dx}\psi(x) \tag{8}$$

in any cross-section. (The factor 4 is the product of the number of silicon electron valleys participating in the transport by the spin degeneracy.)

Equation (5) neglects electron scattering inside the channel, i.e. describes ballistic electron transport. This assumption is sometimes questioned on the basis of experiments with doped silicon channels, that indicate considerable scattering. However, results of recent experiments [62] indicate that electron mobility in undoped SOI layers as thin as 5 nm may be very high (> 400 cm^2/V-s), i.e. essentially the same as in good bulk MOSFETs [63] in the equivalent perpendicular field (pressing the electron wavefunction "centroid" equally close to the Si/SiO_2 interface). If the similar surface roughness can be sustained down to $t \sim 2$ nm, the mobility should be about 200 cm^2/V-s (decreasing by a factor of two because of electron scattering at both interfaces). This mobility (μ) corresponds to an elastic scattering time $\tau = m_l\mu/e$ close to 25 fs. (The inelastic relaxation time for relevant energies is much longer, of the order of 100 fs [64].) For the most important transport region, where the effective potential $U(x)$ is close to its maximum, the electron kinetic energy is of the order of thermal energy $k_BT \sim 25$ meV, i.e. their average speed $v \approx (3k_BT/m_l)^{1/2}$ is close to 2.5×10^7 cm/s; for this speed the elastic mean free path $l = \tau v$ should be about 6 nm. (This estimate is consistent with results of the recent direct measurements of l in bulk

MOSFETs [65].) While this value of l may be lower than the total channel length L, it is still larger than that of the transport bottleneck, where electrons overcome the potential barrier maximum. This estimate means that the ballistic transport may be a reasonable approximation for MOSFETs with high-quality ultrathin undoped channels.

Equations (5ab)-(7) should be solved with appropriate boundary conditions. For the model shown in Fig. 1, an electron leaving the channel has much more chances to be scattered into the bulk drain than back into the channel [66]. As a result, one can use "completely absorbing" boundary conditions that neglect the backscattering completely. (These conditions are frequently used for the analysis of transistor with thin source and drain, but here this assumption is much more questionable, because backscattering may be quite substantial [67].) In order to make these conditions self-consistent, one should assume that occupation of each particular mode of electron propagation in the channel, described by parameters k_y and E_x, is equal to the equilibrium Fermi function of energy $E = E_x + \hbar^2 k_y^2/2m_l + E_z$ in the source (for the few electrons traveling is the back direction, in the drain). This function depends on the level of doping of source and drain. If device-to-device fluctuation of transistor parameters has to be relatively small (this is necessary for acceptable yield of integrated circuit fabrication), the average number N of dopants in electrode regions immediately adjacent to the channel has to be is much larger than one. For sub-10-nm devices the volume V of these regions is of the order of 0.3×10^{-18} cm^3; hence to keep device-to-device fluctuations below 10% the doping rate N/V should be at least as high as 3×10^{20} cm^{-3}, corresponding to deeply degenerate silicon. The numerical results shown below correspond to this value, since higher doping degrades the MOSFET performance.

2.3. *Results*

Figures 2-4 show typical results of the numerical solutions of the equations of the model discussed above. One can see that for relatively long devices ($L = 10$ nm) the characteristics are close to ideal: at positive gate voltage the current rapidly saturates at a level considerably larger than the industrial standard (for n-MOSFETs, 600 μA/μm, i.e. 6 A/cm [1]), while the subthreshold curve slope is close to the perfect, thermally-determined value 60 mV/decade. However, as soon as the gate length L is reduced below approximately 5 nm (channel length L_c, below $\sim$ 8 nm), transistor performance starts to degrade. In particular, the saturation becomes less pronounced and is achieved at higher source-drain voltage V. The subthreshold curve slope (Fig. 3) becomes considerably lower than the perfect and shows increasing dependence of V (the so-called drain-induced barrier lowering, or "DIBL") as L decreases. Moreover, the subthreshold curves (plotted on semi-log scale) start bending upward for large negative values of V_G, due to the contribution from quantum-mechanical source-to-drain tunneling along the channel.

The overall degradation of the transistor can be characterized by voltage gain, i.e. the derivative $G_V = \partial V/\partial V_G$ taken at a fixed drain current J. In good MOSFETs, $G_V \to \infty$ at saturation, so this is not a very popular engineering figure-of-merit. (In the sub-threshold region, the notion of DIBL that is essentially G_V^{-1}, is used instead.) However, as the transistor degrades, the voltage gain becomes an important characteristic, since digital logic circuits fundamentally require $G_V > 1$ for their operation. Figure 4 shows G_V as a function of the gate voltage V_G, at fixed drain-source voltage $V = 0.5$ V. One can see that as the gate length is reduced, G_V rapidly decreases, its maximum approaching unity at $L \approx 2$ nm, i.e. at the channel length $L_c \equiv L + 2t_{ox}$ about 5 nm.

2.4. *Discussion*

Detailed analysis of the results presented above shows that two effects contribute comparably to the device degradation at $L \to 0$: a loss of electrostatic control of the bottleneck potential U_{max} by the gate voltage V_G, and source-to-drain tunneling along the channel.

In order to estimate the electrostatic degradation analytically, one can use the 2D Laplace equation to find electrostatic field distribution for a

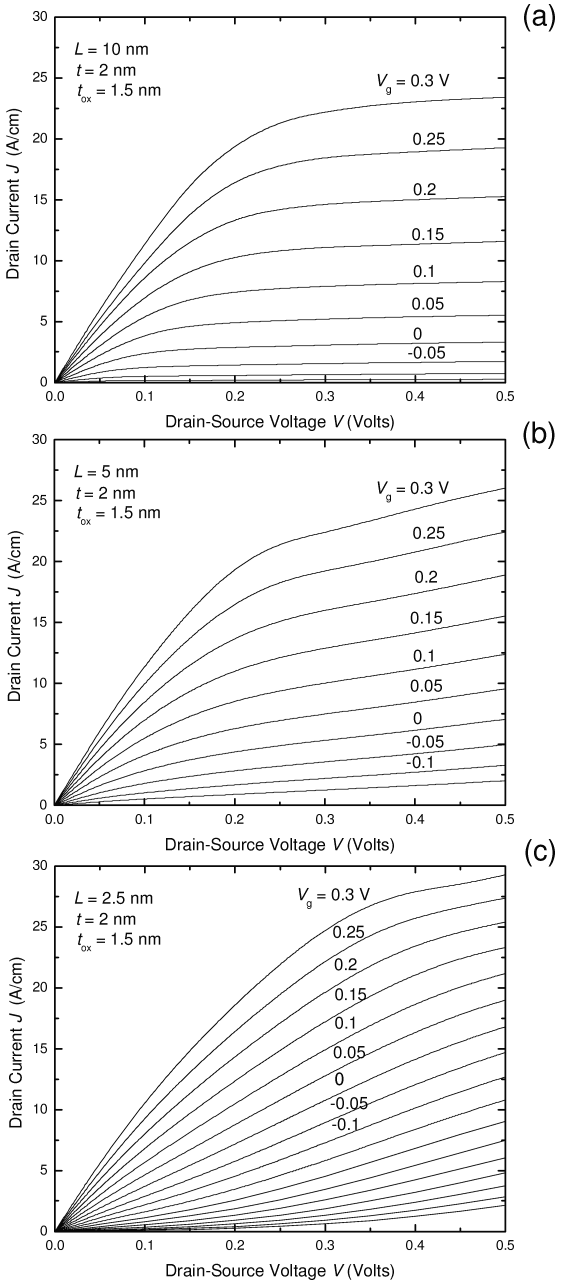

Fig. 2. Drain $I - V$ characteristics of double-gate n-MOSFETs for 3 values of gate length L, numerically calculated [59, 61] using the model shown in Fig. 1. Here and in Figs. 3, 4, 6 and 7 below, the electrode doping level is $3{\times}10^{20}$ cm^{-3}.

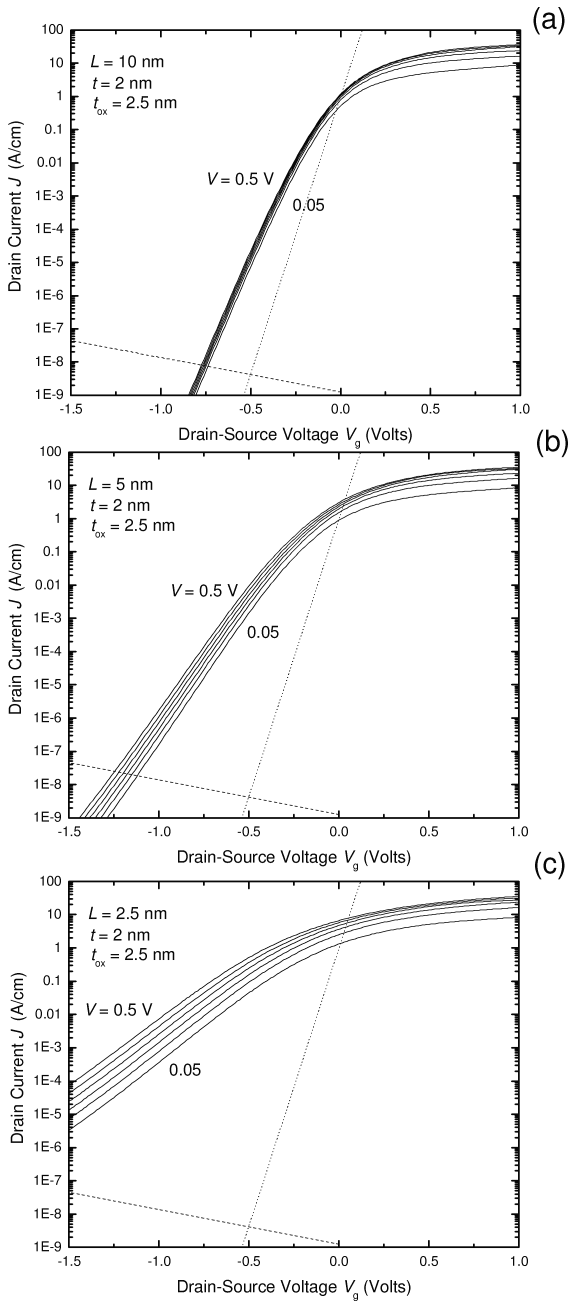

Fig. 3. Subthreshold curves of double-gate n-MOSFETs (Fig. 1) for 3 values of gate length L, each for 10 values of drain-source voltage V (with 50-mV steps) [61]. The dashed lines show the estimated gate oxide leakage current. Notice that the gate oxide is thicker than in Fig. 2. (Subthreshold curves are mostly important for memory applications where oxide leakage should be small). Dotted lines show the ideal slope of 60 mV/decade.

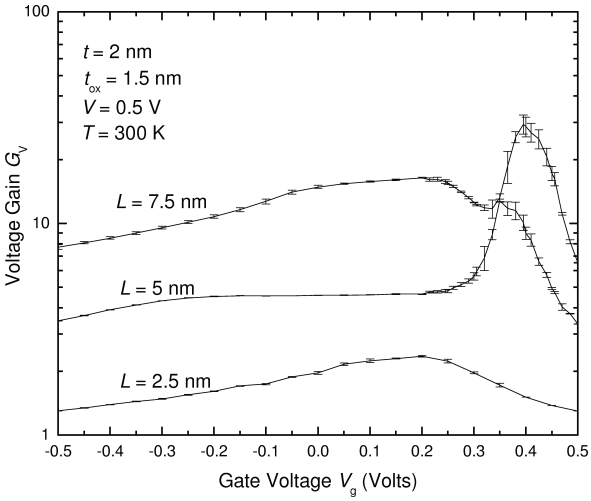

Fig. 4. Voltage gain $G_V \equiv \partial V/\partial V_{G|J=const}$ of nanoscale MOSFETs as a function of gate voltage for $V = 0.5$ V and several values of gate length L [59, 61].

simplified model of the device [43] in the depletion regime ($n = 0$). The assumption of exponential dependence of the electrostatic potential on x, $\Phi(x, z) = \Phi(z)\exp(\pm x/\lambda_E)$, readily yields the following equation for λ_E:

$$(\varepsilon/\varepsilon_{ox})\tan(t/2\lambda_E)\tan(t_{ox}/\lambda_E) = 1. \qquad (9)$$

Figure 5 shows levels of constant λ_E on the [t, t_{ox}] plane, calculated using this equation. It is evident that λ_E decreases with both t and t_{ox}. For small values of the ε_{ox} /ε ratio, and close values of t and t_{ox}, the following relations are approximately satisfied:

$$(\varepsilon_{ox}/\varepsilon)^2 t_{ox}/\varepsilon_{ox} \ll t/2\varepsilon \ll t_{ox}/\varepsilon_{ox}. \qquad (10)$$

In this case, Eq. (9) is reduced to a simple expression,

$$\lambda_E = (\varepsilon t t_{ox}/2\varepsilon_{ox})^{1/2}, \qquad (11)$$

following also from the so-called "parabolic approximation" [68, 69]. Equation (11) allows a very simple interpretation: this is the standard length $(C_s/C_p)^{1/2}$ of field penetration into a 1D line of series capacitances $C_s = t\varepsilon\varepsilon_0$ (representing the longitudinal capacitance of silicon channel per unit length and width) shunted by parallel capacitances $C_p = 2\varepsilon_0\varepsilon_{ox}/t_{ox}$ (the specific transversal capacitance of two oxide layers in parallel).

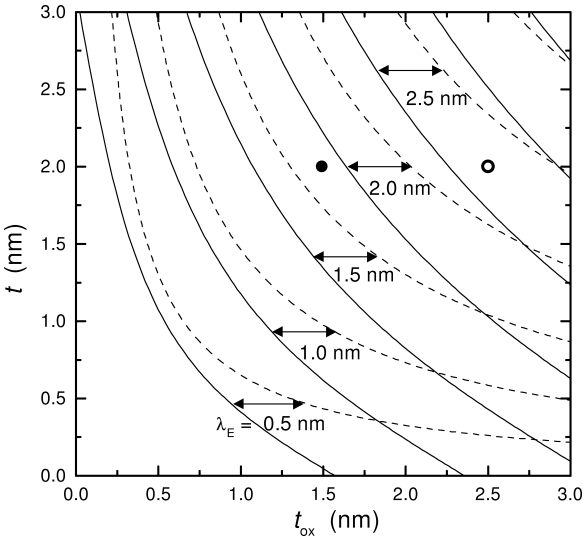

Fig. 5. The contours of fixed characteristic electrostatic length λ_E on the plane of channel thickness t and oxide thickness t_{ox} for Si double-gate MOSFETs ($\varepsilon = 12$, $\varepsilon_{ox} = 3.9$). Solid lines: Eq. (9); dashed lines: the parabolic approximation given by Eq. (11). Solid and open points show the parameter sets used for Fig. 2 ($t_{ox} = 1.5$ nm) and Fig. 3 and 4 ($t_{ox} = 2.5$ nm), respectively.

One can expect the transistor electrostatics to degrade seriously if the channel length is reduced below $\sim \pi\lambda_E$. For the parameters accepted for Fig. 3 ($t = 2$ nm, $t_{ox} = 1.5$ nm, see the solid point in Fig. 5) this gives the channel length limitation $L_c \approx 5.5$ nm, i.e. $L \approx 2$ nm (Fig. 1), while the increase of t_{ox} to 2.5 nm (the open point in Fig. 5) increases the minimum L_c to ~8 nm, i.e. L to about 3 nm. These estimates are in a reasonable agreement with numerical results shown in Fig. 2 and 3, so that one may use Fig. 5 to estimate the transistor electrostatics degradation for other parameters. It is commonly assumed that SiO_2 gate oxide with acceptable leakage and reliability [70-73], and SOI layers [74, 75] with acceptable roughness may be both ultimately thinned to 1 nm. In this case Fig. 5 shows that the electrostatic limit $\pi\lambda_E$ on the channel length is around 3 nm. Some further (probably, modest) improvement of electrostatics can apparently be achieved using the thinned source and drain and more complex structures with nonuniform ("graded") channel and/or gates – see, e.g., Refs. 76-80. Finally, the use of new, high-ε dielectrics such as ultradense silicon oxyni-

tride [81], hafnium oxide [37, 82], zirconium silicate [83], or aluminum oxide [84] (see also Chapter 4 of this collection) may allow the channel length limit to be pushed down a little bit further, possibly to about 2 nm.

However, all these efforts may be inadequate, because of the second important limitation on L_c imposed by source-to-drain tunneling. Let us give a simple estimate of this effect. Since for very short MOSFETs (with channel length of the order of $\pi\lambda_E$) the potential distribution along the channel is rather smooth, a reasonable estimate of tunneling importance may be obtained from the famous Kemble formula (see, e.g., Ref. 85):

$$D(E_x) = \frac{1}{1+\exp\left[2\pi\frac{U_{\max}-E_x}{\hbar\omega}\right]}, \qquad (12)$$

for the WKB transparency of electron tunneling under an inverted quadratic potential $U(x) = U_{max} - m_l\omega^2(x - x_0)^2/2$. By coincidence, this dependence of barrier transparency on energy E_x has exactly the same functional form as the Fermi distribution of the incident electrons. Using this fact, it is straightforward to show that quantum tunneling under such a barrier dominates over "thermionic" charge transfer over the barrier if the physical temperature T is lower than the so-called "inversion temperature"

$$T_{inv} \equiv \hbar\omega/2\pi k_B. \qquad (13)$$

For the effective potential $U(x) = E_z - e\langle\Phi\rangle$ equal to the Fermi level ε_F of source and drain at $(x-x_0) = \pm L_c/2$, and peaking at U_{max} over that level, we obtain that tunneling dominates at $L_c < L_T$, where

$$L_T \equiv (\hbar/\pi k_B T)(2U_{max}/m_l)^{1/2}. \qquad (14)$$

For $U_{max} = 0.05$ eV (which is the typical barrier height in the transistors discussed above) and $T = 300$ K, Eq. (14) yields $L_T \approx 7$ nm. This estimate agrees well with the numerical results shown in Fig 2-4. Shorter devices, like those shown in Figs. 2c and 3c, operate essentially as "tunnel transistors" were gate voltage controls electron tunneling through the barrier. [1] For devices with channel length $L_c \lesssim L_T$, this control still may be effective, tunneling corresponding crudely to an increase of effective temperature from T to $T_{inv} \propto 1/L_c$. This conclusion is (at least qualitatively) confirmed by first experiments [7, 8, 15, 16] with sub-10-nm transistors.

The control is, however, virtually lost as soon as $k_B T_{inv}$ becomes comparable to U_{max}, because the tunnel barrier becomes almost completely transparent. Formula (13) shows that this happens at

$$L_c \approx L_{min} = \hbar/(2m_l U_{max})^{1/2}. \qquad (15)$$

U_{max} can hardly be larger than half the bandgap (otherwise Zener tunneling [2] begins), giving for silicon $L_{min} \approx 2$ nm. Notice, however, that in ultrathin channels the conduction band edge rises by E_z given by Eq. (4), while the valence band edge lowers according to a similar formula, but with the hole rather than electron mass. As a result, the bandgap grows (by as much as ~0.5 eV at $t = 1$ nm), so as a matter of principle the length limitation due to source-to-drain tunneling may be decreased even a little bit further.

2.5. *Parameter sensitivity*

The theoretical predictions made above seem very optimistic: to summarize, they indicate that physics allows FET channel length to be scaled down as deeply as to at least 2 nm, still enabling the performance necessary for operation of logic and memory circuits. However, these results also indicate two major challenges on the way to approaching these limits in commercial practice.

The first problem the rapidly increasing power dissipation [49, 59]. For relatively long transistors ($L_c \gg l$) that are well described by the drift-diffusion model, the specific power P_0 per unit

[1] Earlier, tunnel transistors were discussed mostly in the context of structures with metallic or silicide source and drain which create additional Schottky barriers at channel interfaces – see, e.g., Sec. 9.7 of Ref. 2 as well as recent publications [87-89] and references therein. These barriers cause additional reduction of device transparency; as a result, in most cases the tunnel transistor transconductance and ON current were rather low.

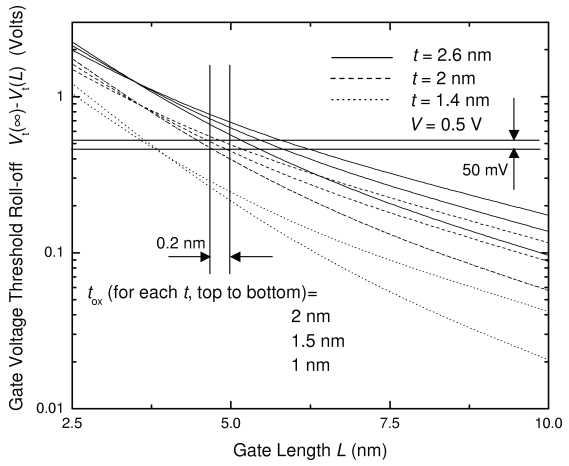

Fig. 6. Threshold voltage shift (relative to that for $L \to \infty$) as a function of gate length L, for all combinations of 3 values for oxide thickness t_{ox} and 3 values for channel thickness t [61].

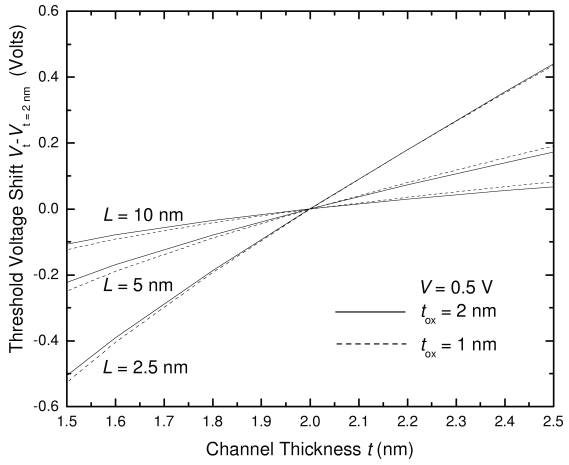

Fig. 7. Threshold voltage dependence on the channel thickness t[61].

channel width decreases with L_c, because drawing a certain ON current density requires approximately the same longitudinal electric field E, but lower drain voltage $V \sim EL_c$. (The real relation of P_0 and V is more complex and depends on many details [3-6], but this does not affect the result qualitatively.) However, as soon as L_c becomes comparable with the mean free path, the power decrease stalls, because the ON current is now limited by the source electron supply exhaustion [39, 40] rather than channel mobility. Moreover, at $L_c \lesssim 10$ nm the transistor degradation due to electrostatics and source-to-drain tunneling leads to increase of P_0 with the channel length reduction, even at the optimum choice of the power supply voltage V_{DD} [49, 59]. This increase would further exacerbate the power management problem that is very severe even now [1, 90, 91].

However, even more significant is the second problem, the rapidly increasing sensitivity of transistor characteristics to unavoidable random variation of geometrical parameters, due to fabrication uncertainties. For example, Fig. 6 shows the decrease ("roll-off") of the threshold gate voltage V_t (defined as the value of V_G providing a certain small drain current) with a reduction of the gate length. The plot shows that small changes in the length definition may lead to large variations of V_t. These variations should be compared with the minimum power supply voltage V_{DD} necessary for driving the device.

For example, consider a relatively long device with physical gate length $L = 9$ nm, planned by the ITRS [1] for commercial introduction in approximately the year 2016. The same document (see, e.g., Table 57) predicts that by then the critical dimension control accuracy (at the 3σ level) will reach $\sim$0.7 nm. Figure 6 shows that this control would enable variations of V_t to be kept within approximately 50 mV. On the other hand, Fig. 2a shows that in order to keep ON current density at the standard 600 μA/μm level [1], the voltage swing V_{DD} should be above 300 mV; hence the 50 mV variation seems acceptable.

Now consider a shorter device with $L = 5$ nm, t = 2 nm, and t_{ox}= 1.5 nm ($L_c = 8$ nm). According to Fig. 6, in order to keep fluctuations of V_t below 50 mV, the critical dimension should be controlled better than $\sim$ 0.2 nm, much tighter than the farthest ITRS projections for even the most advanced lithographic techniques such as EUV [97, 98], currently in the stage of laboratory development.

Moreover, Fig. 7 shows that V_t is even more sensitive to the channel thickness t, mostly because of the strong dependence of the quantum confinement energy E_z on t – see Eq. (4). For example, in order to keep fluctuations of V_t at the same level (below 50 mV), t should be controlled better than $\sim$0.1 nm, a very hard task indeed – see, e.g., Chap-

ter 3 of this collection.

The necessity to ensure such tight control of device dimensions will lead to a rapid increase of fabrication cost facilities, that may reach the point of diminishing economic returns. As a result, the Si-MOSFET-based exponential Moore-Law progress may stop at $L \sim 10$ nm, i.e., long before fundamental physical limits have been reached.

2.6. *Alternative FET materials*

In view of this situation, it is important to take a second look at the numerous suggestions to replace silicon with another channel material. Among those, single-wall carbon nanotube FETs [92-96] seem to be most promising. (Results were obtained for multi-wall carbon nanotubes [96] are less impressive.) A possible advantage of these structures could be smaller interface roughness that may allow to implement small effective values of channel thickness t ($\sim$1.5 nm) while still sustaining reasonable mobility. However, for sub-10-nm transistors the requirement of very high mobility fades away, because current is more limited by source exhaustion. (Since electrode material is different from that of the nanotube, current may be also substantially limited by contact resistances.)

In addition, present day production methods give nanotubes with random helicity, which strongly affects the parameters of longitudinal electron transport, effectively changing them from metals all the way to semiconductors with considerable bandgap. Finally, carbon nanotubes (or any other channel material) cannot help solve the key problem of transistor parameter sensitivity to channel length – see Sec. E above. As a result, I do not believe that an alternative channel material may be a universal remedy against the sub-10-nm transistor woes.

3. Single Electron Devices

3.1. *Motivation*

The problem outlined in the end of the last section motivates a search for new nanoscale electronic devices based on different physics. General guidelines for such a search may be provided by the following arguments [99] based on the famous particle-wave duality. Quantum mechanics says that electrons may behave either as discrete particles or continuous de Broglie waves, depending on experimental conditions. (Surprisingly enough, a sufficiently clear understanding of these conditions for conduction electrons in solids was achieved not so long ago, in the 1980s – for reviews see, e.g., Refs. 100-102.)

Consider for example a generic situation where two parts of a conductor are separated by some interface, and ask whether the electric charge of each conductor is a multiple of the fundamental charge e at any instant (this is natural for the particle picture) or may be continuous (the wave picture allows this, because the wavefunction of each electron may be split between the two parts). The answer to this question turns out to be dependent on whether the effective tunnel resistance R of the interface is larger or smaller than the natural quantum unit of resistance

$$R_Q \equiv \hbar/e^2 \approx 4.1k\Omega. \tag{16}$$

If the resistance is low, $R \ll R_Q$, the charge of each conductor may be continuous, but in the opposite limit it may be only a multiple of e.

This relation may be derived and explained in numerous ways; perhaps the simplest interpretation is as follows. In a closed ("Hamiltonian") quantum system, the characteristic energy of quantum fluctuations per degree of freedom is $E_Q \sim \hbar\omega/2$, where ω is a characteristic frequency. In contrast, each part of the conductor we are discussing, concerning its electric charge degree of freedom, is an "open" system, strongly interacting with its environment (in classics, corresponding to an RC relaxator, rather than an LC oscillator). For such a system, $\hbar\omega$ should be replaced by $\hbar/\tau$, with $\tau = RC$, where C is the capacitance between the two conducting parts. Transfer of a single electron between the parts causes an electrostatic energy change of the order of $E_C \sim e^2/2C$. Comparing E_C and E_Q we notice that C cancels, and see that if $R \ll R_Q$, quantum fluctuations smear out the electrostatic energy difference which tries to keep the electric charge of each part constant.

To comprehend the importance of this result, let us combine it with the so-called Landauer formula for the interface conductance $G \equiv 1/R$ [100-102]:

$$G = (e^2/\pi\hbar)\Sigma_i D_i \qquad (17)$$

where D_i is the interface transparency (i.e. the probability of electron transmission) for a particular transversal mode of electron propagation; the sum is over all the modes. Comparison of Eqs. (16) and (17) shows that if the conductor cross-section is so narrow that quantum confinement makes only one propagating mode possible, the condition of electric charge discreteness takes a simple and natural form: $D \ll 1$. However, for devices with a larger cross-section A, the restriction on the average transparency is much more severe: $D \ll 1/N$, where N is the transversal mode number. (For a degenerate conductor, N is of the order of λ_F^2/A where λ_F is the Fermi wavelength, typically of the order of 1 nm.)

An FET transistor is a good example of a device where the number n of electrons in the channel is never quantized, because the boundaries between the electrodes and channel are typically highly transparent ($D \sim 1$). As a result, these transistors do not exhibit single-electron charging effects even if the *average* number of electrons in the channel is small. An adequate understanding of such devices may be achieved using the wave language, used in particular in Sec. 2 above.

Theoretically, for nanoscale "wave" devices ($R \ll R_Q$), the FET-type control of transport is not the only possible mode of operation: the effects of quantum interference of electron de Broglie waves can be, as a matter of principle, used for this purpose as well. In the 1980s and early 1990s, much attention was focused on such "quantum electronic devices" – see, e.g., Refs. 103-105. However, later the prospects for their practical applications have been recognized as rather poor, mostly for the following reason. In contrast to optical phonons (which obey Bose statistics), charge carriers in solids are fermions and in particular obey the Pauli principle: each of them must have a different energy, and hence a different de Broglie wavelength. Hence, high-contrast interference patterns require operation with either a single transversal mode or a small number of modes. The accuracy δL of size definition of the nanostructures that single out such a mode from a continuum, and then handle its interference, should be much better than the de Broglie wavelength $\lambda = h/(2mE)^{1/2}$ of the used electrons. Simultaneously, the electron energy E should be well above the thermal fluctuation scale (typically, $\sim k_B T$) in order to avoid interference pattern smearing by thermal fluctuations. Combining these two requirements, and plugging in fundamental constants, we may see that for room temperature δL should be, as in nanoscale MOSFETs, well below 1 nm. (This is only natural, because λ has the same order of magnitude as L_T defined by Eq. (14)). This simple estimate show that quantum interference devices do not have any substantial advantage over FETs for sub-10-nm scaling. [2]

The above discussion pertains to *spatial* quantum interference; one can also consider using *temporal* quantum coherence of electrons for information processing. The most prominent ideas put forward in this field are those of quantum encryption and quantum computing – see, e.g., [113]. Presently, the former goal seems to be much closer than the latter; however, both of them represent rather narrow application niches. For most tasks faced by digital electronics, quantum computing does not seem to offer significant advantages over

[2] I am aware of two important exceptions of this conclusion. First, if the only critical dimension of a device may be defined by film thickness, it may be readily controlled with sub-nm accuracy. This is the case of vertical resonance diodes [106, 107]. If such devices with acceptably large peak-to-valley ratio are implemented in CMOS-compatible technology (see, e.g., Ref. 108) there will be hope for their practical introduction. Unfortunately, the range of possible applications of these two-terminal devices is probably limited to fast semiconductor memories (challenging the current SRAM technology). The second important exception are Cooper pairs in superconductors, that are bosons rather than fermions. As a result, many Cooper pairs may have exactly the same wavefunction, enabling the so-called macroscopic quantum interference effects in structures much larger than λ - see, e.g., Refs. 109, 110. These effects are used, in particular, in fast, ultra-low-power Rapid Single-Flux-Quantum (RSFQ) logic circuits – see, e.g., recent reviews [111, 112]. Unfortunately, these circuits need deep refrigeration, the fact that has so far has hindered their wide practical introduction.

the usual ("classical") computing, while being much harder to implement. Because of this reason, I will abstain from discussing this issue in this chapter which deals with potentially practicable digital technologies. (A brief review of quantum computing may be found in the last chapter of this collection.)

To summarize, in the category of "wave" electron devices ($R \ll R_Q$) we are left with not much more than the field effect transistors. This motivates us to move to nanoscale structures with high impedance ($R \gg R_Q$), dominated by single-electron charging effects.

3.2. *Single-electron box*

By now, the basic physics of single-electron devices has been developed quite deeply, but is known within a much narrower circle than that of MOSFETs. Therefore I will give its brief review. (More detailed reviews may be found, e.g., in Refs. 114-117.) Let us start with a generic single-electron device, frequently called the "single-electron box" - Fig. 8a. [3]

The device consists of just one small conductor ("island") separated from an external electrode by a tunnel barrier with high resistance,

$$R \gg R_Q. \tag{18}$$

An external electric field may be applied to the island using a capacitively coupled gate electrode. The field changes the local Fermi level of the island and thus determines the conditions of electron tunneling. Elementary electrostatics shows that the energy of the system may be presented as

$$W = Q^2/2C_\Sigma - (C_g/C_\Sigma)QV_g + const, \tag{19}$$

where $Q = -ne$ is the island charge (n is the number of uncompensated electrons), C_g is the island-gate capacitance, while C_Σ is the total capacitance of the island (including C_g). It is convenient to rewrite Eq. (19) expression in another form,

$$W = (Q_0 - ne)^2/2C_\Sigma + const, \quad Q_0 \equiv C_g V_g, \tag{20}$$

where parameter Q_0 is usually called the "external charge". From its definition, it is evident that in contrast with the discrete total charge Q of the island, the variable Q_0 is continuous, and may be a fraction of the elementary charge e.

At sufficiently low temperatures,

$$k_B T \ll E_C, \quad E_C \equiv e^2/C_\Sigma, \tag{21}$$

the stationary number n of electrons in the island corresponds to the minimum of W; an elementary calculation using Eq. (20) shows that Q is a step-like function of Q_0, i.e. of the gate voltage (Fig. 8b), jumping by e when

$$Q_0 = e(n + \frac{1}{2}), \quad n = 0, \pm 1, \pm 2, \ldots \tag{22}$$

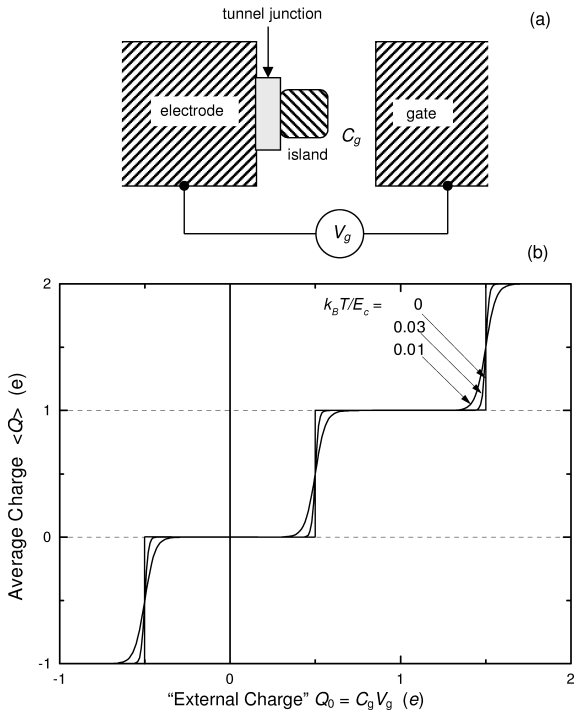

Fig. 8. Single-electron box: (a) schematics, and (b) the "Coulomb staircase", i.e. the step-like dependence of the average charge Q of the island on the gate voltage V_g, for several values of relative temperature.

[3] The basic physics of this device was understood by Lambe and Jaklevic [118] as early as in 1969, on the basis of their experiments with disordered granular structures, while the first quantitative theory of the box was developed by Kulik and Shekhter [119]. The first experiments with an individual box were, however, carried out much later [120], after key experiments with other, more complex, single-electron devices.

If the temperature is increased to $k_BT \sim E_c$, the system has non-vanishing probability p_n to be in other states as well, with

$$p_n = \frac{\exp\{-W(n)/k_BT\}}{\sum_n \exp\{-W(n)/k_BT\}}. \qquad (23)$$

A straightforward calculation of the average charge $\langle Q \rangle = -\sum_n enp_n$ yields the pattern shown in Fig. 8b: the step-like dependence of charge on gate voltage is gradually smeared out by thermal fluctuations. This is typical for all single-electron devices, so that the operation temperature of most of them should satisfy Eq. (21). (A notable exception are single-electron temperature standards that operate at $k_BT \sim E_C$ – see, e.g., Refs. 121, 122.)

The physics of this "Coulomb staircase" is very simple: increasing gate voltage V_g tries to attract more and more electrons to the island. The discreteness of electron charge, provided by low-transparency barriers (with $R \gg R_Q$) ensures that the changes may be only discrete. Notice that in this sense the box (and all other single-electron devices) are not really a "quantum electron device", or at least much less so than the usual field-effect transistor.

3.3. *Single-Electron transistor*

A simple modification of the single-electron box, splitting its electrode into two parts (source and drain), so that voltage V may be applied between them, turns it into a very important device, the single-electron transistor (Fig. 9a). This device, that was first suggested in 1985 [123, 124] and first implemented two years later [125], is clearly reminiscent of an FET, but with a small conducting island limited by two tunnel barriers, instead of the usual channel, connecting the source and drain.

Figure 9b shows a typical set of dc $I-V$ curves of such transistor, for several values of the "external charge" Q_0, defined in the same way as in the single-electron box – see the second of Eqs. (20). One can see that at small drain-to-source voltage V, there is virtually no current, besides the special values of Q_0 given by Eq. (22). The physics of this phenomenon (the "Coulomb blockade") is easy to

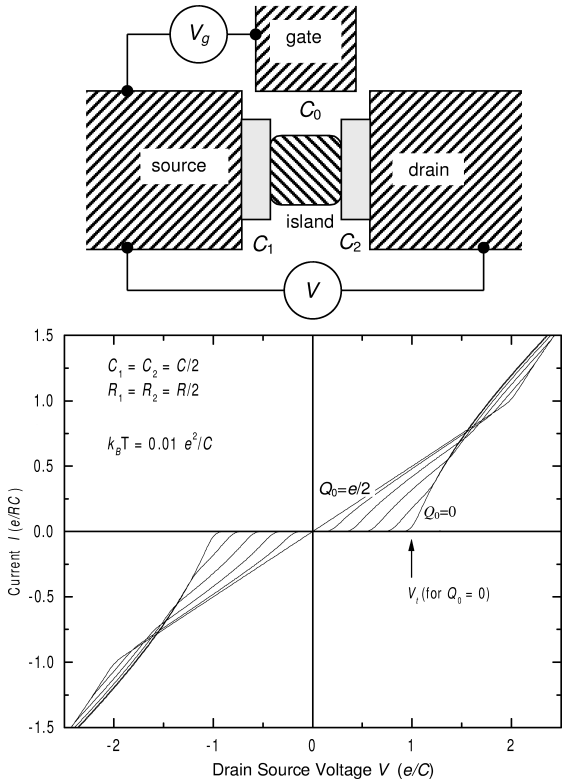

Fig. 9. Single-electron transistor: (a) schematics, and (b) a typical set of source-drain $I-V$ curves of a symmetric transistor for several values of the "external charge" Q_0, i.e. of the gate voltage V_g, calculated using the "orthodox" theory of single-electron tunneling.

understand: even if $V > 0$, and thus it is energy-advantageous for a electron to go from source to drain, on its way the electron has to tunnel into the island first, and change its charge Q it by $\Delta Q = -e$. Such charging would increase the electrostatic energy W of the system

$$W = (Q_0 - ne)^2/2C_\Sigma - eV(n_1C_2 + n_2C_1)/C_\Sigma + const, \qquad (24)$$
$$C_\Sigma \equiv C_g + C_1 + C_2,$$

(where n_1 and n_2 are the numbers of electrons passed through the tunnel barriers 1 and 2, respectively, so that $n = n_1 - n_2$), and hence at low enough temperatures ($k_BT \ll E_C$) the tunneling rate is exponentially low.

At a certain threshold voltage V_t the Coulomb blockade is overcome, and currents starts to grow

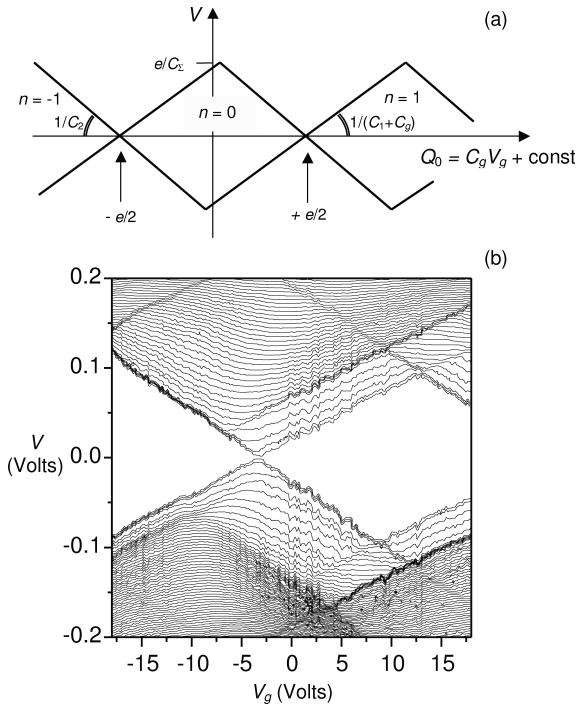

Fig. 10. Coulomb blockade threshold voltage V_t as a function of Q_0: (a) theoretical dependence at $T \to 0$, and (b) experimental contour plots of current on the $[V, V_g]$ plane for an aluminum SET with $E_c \approx 100$ meV at $T = 4.2$ K [126].

with V. The most important property of the single-electron transistor is that V_t is a periodic function of V_g, vanishing in special values of gate voltage, given by Eq. (22) – see Fig. 10. The reason for these so-called "Coulomb blockade oscillations" is evident from the above discussion of the single-electron box: in the special points (22), one electron may be transferred to the island from either drain or source without changing the electrostatic energy of the system even at $V = 0$. Hence, an electron can tunnel from the source to the island and then to the drain even at negligible V , so that $V_t = 0$. As can be readily shown from Eq. (24), at low temperatures the dependence of V_t on V_g is piecewise-linear, with its lower and upper branches forming the so-called "diamond diagram" (Fig. 10) [123].

Since the $I - V$ curves of the transistor are continuous (Fig. 9b), if a small current is fixed by an external circuit, V is close to V_t and also follows the diamond diagram – see, e.g., Fig. 10b. Thus the voltage gain and transconductance of single-electron transistors may change sign depending on the gate voltage – an important difference in comparison with usual field-effect transistors. On the other hand, the same diamond diagram shows that the voltage gain is limited by a capacitance ratio: $(G_V)_{max} = C_g/C_2$ [124]. It may be higher than unity (see, e.g. experiments [127-129]), but hardly much higher than that, especially in room-temperature transistors.

3.4. *Single-electron trap*

Another key device, the "single-electron trap", may also be understood as a generalization of the single-electron box. Let us replace the single tunnel junction in Fig. 8a with a one-dimensional array of $N > 1$ islands separated by tunnel barriers - Fig. 11a. [4] The main new feature of this system is its internal memory, i.e., bi- or multi-stability: within certain ranges of applied gate voltage V_g the system may be in one of two (or more) charged states of its edge island (Fig. 11b).

The reason for this multi-stability stems from the peculiar properties of an electron inside a 1D array: an electron located in one of the islands of the array extends its field to a certain distance [133, 134] and hence may interact with the array edges (is attracted to them). As a result, the electrostatic self-energy of the electron has a maximum in the middle of the array. By applying sufficiently high gate voltage $V_g = V_+$ the energy profile may be tilted enough to drive an electron into the edge island; if the array is not too long, other electrons feel its repulsion and do not follow. If the gate voltage is subsequently decreased to the initial level ($V = 0$), the electron is still trapped in the edge island, behind the energy barrier. In order to remove the electron from the trap, the voltage has to be reduced further, to $V_g < V_- < 0$. As a result, the $n(V_g)$ dependence exhibits regions of bi-

[4] This device was first discussed explicitly in 1991 [130, 131], but in fact it may be considered just a particular operation mode of a more complex device, the single-electron turnstile, invented earlier [132].

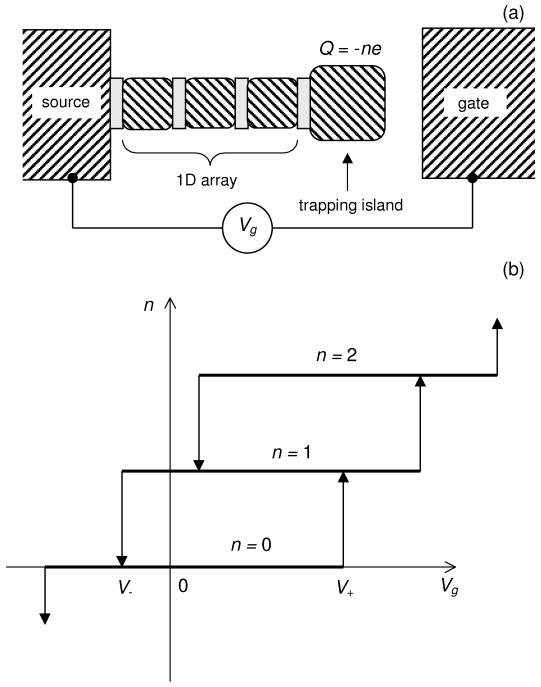

Fig. 11. (a) Single-electron trap and (b) its static characteristic (schematically).

or multi-stability, in which the charge state of the trap depends on its prehistory (Fig. 11b).

The retention time of a certain charge state within the multi-stability region is fundamentally limited by the thermal activation over the energy barrier, and a higher-order quantum process, co-tunneling [135]. The first effect is exponentially low in E_c/k_BT, while the second effect falls exponentially with the array length N. As a result, electron retention time may be very long; experimentally, single electron trapping for more than 12 hours has been demonstrated [136-138].

Note that a similar multi-stability may be also achieved in the simpler device, single-electron box (Fig. 8a) if the tunnel barrier is so thick that the reciprocal tunneling rate Γ^{-1} is longer than the measurement time scale. However, in order to suppress this barrier and ensure fast box recharging, the energy eV_g available from the applied gate voltage should be comparable with the barrier height, typically of the order of a few electron-volts. On the other hand, in the trap the necessary energy is of the order of E_C, and in low temperature experiments may be much lower. However, for room-temperature devices, E_C should be also of the order of 1 eV or higher (see Fig. 13 and its discussion below), and the advantage of the trap over the box fades away.

3.5. *Single-electron parametron*

The last key single-electron device, the so-called parametron [140] is essentially a short segment of a 1D array of islands, galvanically detached from both electrodes. The simplest version of the device uses three small islands separated by two tunnel barriers (Fig. 12a). For simplicity, I will describe its operation for the case when the system is charged by one additional electron, though such pre-charging is in fact unnecessary [141].

Let the parametron be biased by a periodic "clock" electric field $\boldsymbol{E}_c(t)$, oriented vertically, for example, by a slight vertical shift of the central island. (Of course, the same effect may be achieved in a strictly linear array, using a special gate located closer to the central island [141].) This field keeps an extra electron in the central island during a part of the clock period. At some instant, the field E_c reaches a certain value E_t at which electron transfer to one of the edge islands becomes energy advantageous. If the system were completely symmetric, the choice between the two edge islands would be random, i.e. the system would undergo what is called the "spontaneous symmetry breaking". However, even a small additional field $\boldsymbol{E}_s$ applied by a similar neighboring device(s) may determine the direction of electron tunneling at the decision-making moment. Once the energy barrier created by the further change of the clock field has become large enough, the electron is essentially trapped in one of the edge islands, and the field $\boldsymbol{E}_s$ may be turned off. Now the device itself may serve as a source of the dipole signal field $\boldsymbol{E}_s$ for the neighboring cells. The sign of this field (i.e. of the electric dipole moment of the device) presents one bit of information.

Figure 12b shows the phase diagram of the parametron [141], that gives a quantitative description of the properties described above. Each of "ON" states (with the extra electron in one

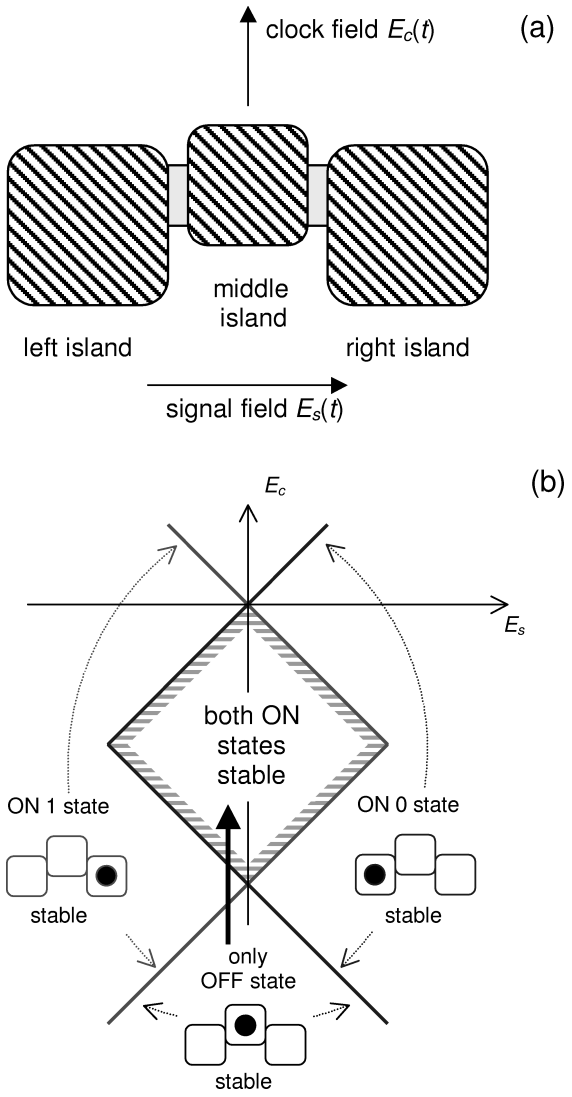

Fig. 12. (a) Single-electron parametron and (b) its phase diagram for the case of precharging by one extra electron. The bold arrow shows the evolution described in the text.

of the edge islands) is stable within an angle-limited region. These regions overlap providing the (hatched) diamond-shaped region of cell bistability. The route from "OFF" state (with the electron in the central island) to one of the ON states, that was described above, in shown by the bold arrow.

Recently, low-temperature prototypes of the single-electron parametron were experimentally demonstrated by two groups [142, 244, 309], albeit one of them prefers to use a different name for this device.

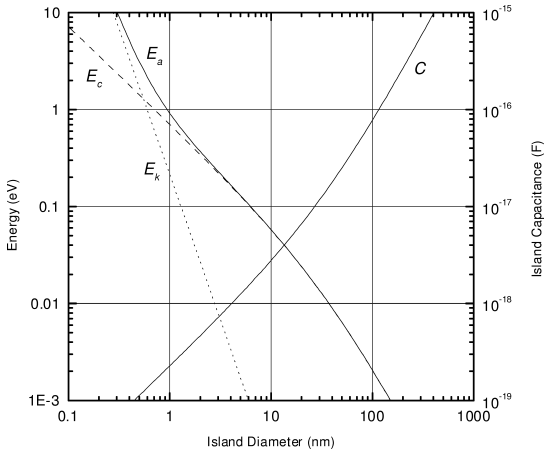

Fig. 13. Single-electron addition energy E_a (solid line), and its components: charging energy E_C (dashed line) and electron kinetic energy E_k (dotted line), calculated for a simple model of a single-electron island [117].

3.6. *Scaling and implementation*

Before discussing the current and prospective applications of single-electron devices, we should consider their scaling. The most important conditions for operation of such devices are given by Eqs. (18) and (21). The former condition, $R \gg \hbar/e^2 \approx 10$ kΩ, is relatively easy to satisfy experimentally, since resistance R of a tunnel barrier grows exponentially with its thickness. On the other hand, this relation shows that the output resistance of single-electron devices is unavoidably high, practically above ~ 100 kΩ. This feature is frequently cited as a drawback. However, what is really important for applications (e.g., the interconnect recharging speed) it the maximum current density per unit width. For a room-temperature SET with $V \sim V_t \sim E_C/e \sim 1$ Volt, the total current would be below 10 μA. However, since width of such transistor has to be very small, ~ 1 nm (see below), the available current density may be well above 1,000 μA/μm, i.e., even higher than that of standard silicon MOSFETs [1].

The latter condition (21) makes the practical implementation of single-electron transistors operating at room temperature rather problematic. Fig-

ure 13 shows the energy E_a necessary to put an additional electron on a transistor island of a certain radius, calculated within the framework of a simple model [117]. This energy is crudely a sum of the electrostatic contribution $E_C = e^2/C$, where C is the island capacitance, and quantum confinement energy E_k. While E_C dominates for relatively large islands, for 1-nm-scale islands with size comparable with the electron de Broglie wavelength, E_k becomes substantial. [5]

Both theory and experiment show that single-electron tunneling effects (i.e. some current modulation by gate voltage in single-electron transistors) become visible at $E_a \sim 3k_BT$. It means that in order to notice these effects at $T \sim 100$ mK (the temperature routinely reached in standard helium-dilution refrigerators), E_a should be above ~ 25 μeV, corresponding to the island capacitance $C \sim 5\times10^{-15}$ F and island size of the order of 1 micron, with tunnel junction area $\sim 0.1\times0.1$ μm^2. Such dimensions can be routinely reached by several methods, including notably the now-classical technique of metal evaporation from two angles through hanging resist mask (most typically formed by direct e-beam writing on a double-layer resist). This method was used, in particular, for the first experimental demonstration of a stand-alone single-electron device [124]; later this technique, complemented with subsequent island oxidation, has been advanced to increase E_a to ~ 0.1 eV in certain samples – see, e.g., Fig. 10b [126]. This is sufficient to see a slight current modulation even at room temperature.

However, for reliable operation of most digital single-electron devices, the single-electron addition energy should be approximately 100 times larger than k_BT. This means that for room temperature operation, E_a should be as large as ~ 3 eV. According to Fig. 13, this value corresponds to island size about 1 nm. Reproducible fabrication of integrated circuits with features so small presents quite a challenge. Most claims of success in this direction have been based on results from evidently irreproducible structures, for example, arrays of nanoparticles deposited between rather distant source and drain. (By chance, one of these particles may be tunnel-coupled to its neighbors much more weakly than are others, forming a single-electron transistor island, while strongly coupled particles effectively merge, forming its source and drain.) Of course, the parameters of such transistors are unpredictable, and there is no hope of using them in integrated circuits.

A more interesting option is fabrication of discrete transistors with scanning probes, for example by nano-oxidation of metallic films [145, 146] or manipulation with carbon nanotubes [147- 149]. For the former devices, single-electron addition energies as high as 1 Volt have been reached [146], though unfortunately the current was very low (below 10^{-11} A). Moreover, the scanning probe techniques are so slow that there is no hope for their use in the fabrication of circuits of any noticeable integration scale, though for discrete devices this approach may be promising.

Several methods which are closer to standard CMOS technology have also been used to fabricate single-electron transistors, mostly by the oxidation of a thin silicon channel until it breaks into one or several tunnel-coupled islands – see the pioneering work [150] and recent results [151-154]. The highest values of E_a reached in those efforts (~ 250 mV [154]) are still not high enough for logic applications, but for some memory cells (see below) this value is almost sufficient. A problem with this approach is that the parameters of the resulting transistors are still highly irreproducible, and there is little chance of changing this situation without the development of patterning technologies with a sub-nm resolution – a very distant goal indeed (see, e.g., the next chapter of this collection).

A radical way to overcome the reproducibility is to use natural 1-nm-scale objects of exact size and shape: molecules. Starting from the mid-1990s, several groups have managed to measure electron transport through a single molecule (or a few molecules in parallel) placed between two

[5] Notice that at $E_k \gtrsim E_C$, the basic ("orthodox") theory of single-electron tunneling [115] should be modified [142, 143] (see also review [144]), although the device implications do not change significantly.

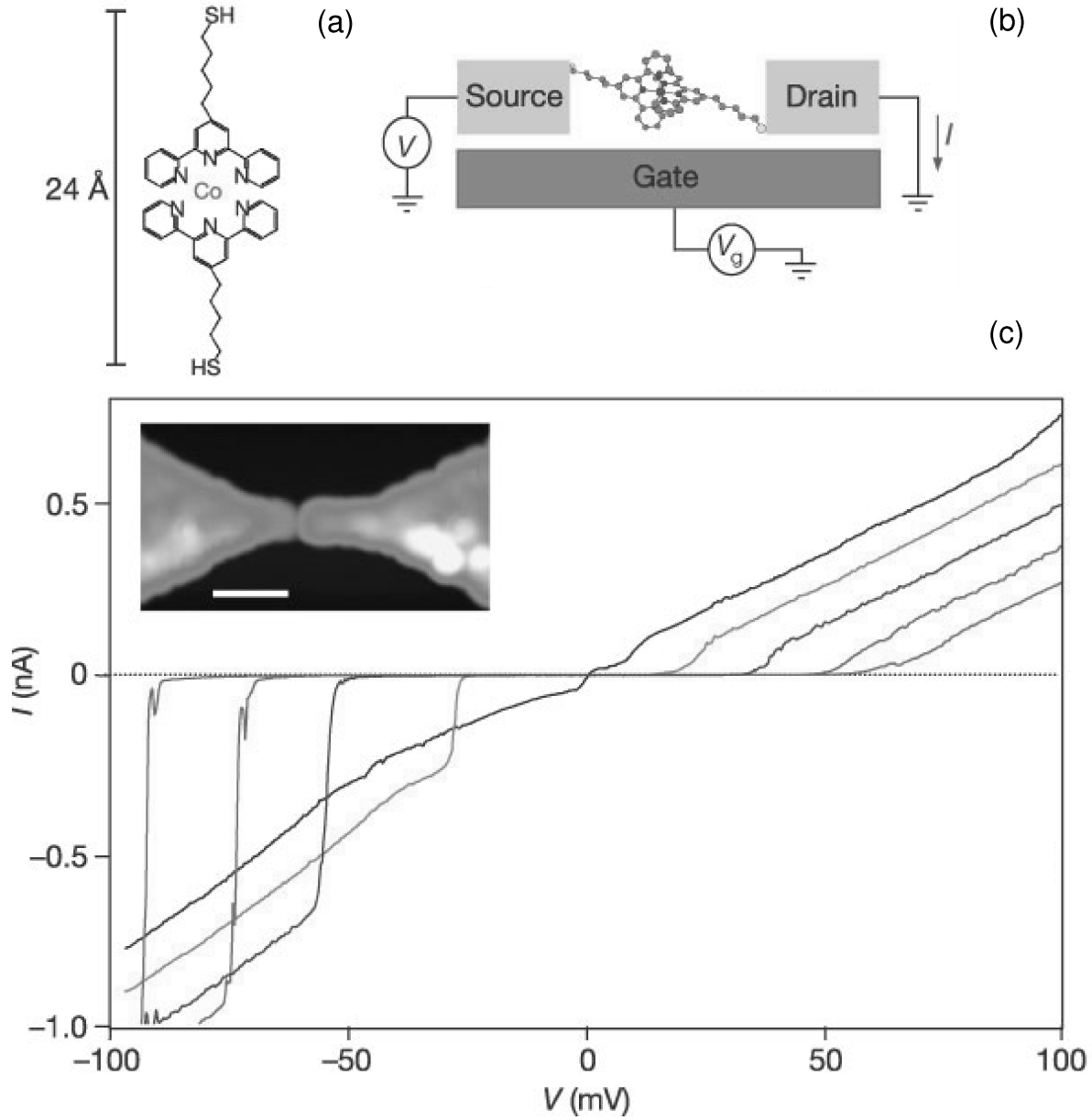

Fig. 14. A single-molecular implementation of single-electron transistor [172]: (a) the molecules that provided the SET-like characteristics shown in panel (c) for several gate voltages. Panel (b) shows the scheme of the experiment, while a SEM picture of the used nanoelectrodes is shown in the inset.

metallic electrodes[6] [155-178, 314, 315]. During this period, the accuracy and reproducibility of such experiments has been improved dramatically, and some of the observed phenomena have already been understood, at least on semi-quantitative basis. The first attempts at qualitative explanation of the transport data [179-184] (see also Chapter 6 of this collection) have been only partly successful so far, apparently because the chemistry of molecule-to-metal interfaces still has to be understood in more detail for each particular method of structure formation.

Besides some very short molecules (see, e.g., Refs. 172, 173, 176) that feature high low-voltage conductance $G \sim e^2/h$ (that is interpreted, in terms of Eq. (17), as high transparency D of the molecule for at least one transversal mode), most samples show $I - V$ curves reminiscent of those shown in Fig. 9b – see, e.g., Fig. 14b. This suppression of current at low voltages is typically interpreted as follows: if the molecule is long enough, it would typically feature a part where electron orbitals are not too much affected by metallic electrodes. These orbitals typically feature considerable discreteness, and hence a substantial energy gap Δ between its higher occupied molecular orbital (HOMO) and lowest unoccupied orbital (LUMO). If the applied voltage is small, Fermi levels of both electrodes have a good chance to be inside this gap, so the conductance is small. (If $\Delta \gg k_B T$, the only mechanism is quantum-mechanical tunneling through the molecule as a whole; this tunneling is very small is the molecule length is above a few nanometers – cf. the discussion of MOSFETs in Sec. 2.3) As the applied voltage increases, either HOMO or LUMO is eventually aligned with one of the Fermi levels, and electrons may pass through the molecule in two hops (from one electrode to the inner orbital and then to another electrode), resulting in considerable increase of current.

This physics is very close to that of single-electron transistor (essentially identical to it, if the discreteness of inner energy levels is taken into account [142, 143]). Consequently, one might expect that if the electrostatic potential of the molecule interior could be changed by voltage V_g applied a galvanically-insulated gate electrode, the Coulomb blockade threshold voltage would be changed in the piece-linear fashion shown in Fig. 10a. This conclusion has been confirmed in first experiments [156, 165, 167, 169, 171-173, 314] where such gate has been arranged – see Fig. 14b. Hence, one may state that single-molecule, single-electron transistors have already been implemented.

So far, these devices have been formed by techniques excluding practical fabrication of integrated circuits. However, there are very good prospects for chemical synthesis of special molecules that would combine the structure suitable for single-electron tunneling with the ability to self-assembly from solution on prefabricated nanostructures, with acceptable yield – see, e.g., the discussion in Ref. 185. Then a way to generically inexpensive fabrication of VLSI circuits would be open. We will come back to this opportunity in Sec. 4 below.

3.7. *Random background charge problem*

Besides fabrication, single-electronics faces one more serious problem. Let a single charged impurity be trapped in the insulating environment, say on the substrate surface, at a distance r from a single-electron island, comparable to its size a. The impurity will polarize the island, creating on its surface a polarization ("image") charge of the order of $e(a/r)$, that is effectively subtracted from the external charge Q_0 – see, e.g., Eq. (20). This charge affects all characteristics of single-electron devices, for example, in the single-electron transistor it shifts the Coulomb blockade threshold V_t – see Fig. 10. For $r \sim a$, this shift may be large, of the order of $(V_t)_{max}$, even from a simple impurity. Using even the most optimistic estimate compatible with experimental data, 10^9 cm^{-2} [186], for the minimum concentration of charged impu-

[6] The list of employed configurations includes scanning tunnel microscopes (STM) [155, 156, 158, 164-166, 168, 169], "break junctions" (mechanically controlled cracks in narrow metallic wires) [156, 160, 162, 170, 174], crossings of two narrow wires [159, 163,], nano-orifices in a thin film [161, 175], gaps in nanowires, either narrowed [167] or created [172, 173, 314] by electromigration, and even junctions fabricated on an STM tip [171].

rities, and assuming 1-nm island size, we can estimate that at least 10^{-3} part of the chip area would be "poisoned", so that the same fraction of single-electron devices will have an unacceptably large background charge fluctuation, $\delta Q_0 \gtrsim 0.1\ e$.[7]

A possible way to circumvent this problem is the use single-electron transistors with resistive (rather than capacitive) coupling [124, 187-188], which are insensitive to background charge. These devices require "Ohmic" resistors with very high resistance (above ~ 1 MΩ) and quasi-continuous ("sub-electron") transfer of charge that would provide the compensation for the fractional part of the random background charge. Such resistors have been indeed demonstrated (see, e.g., Ref. 190), however, their implementation for room temperature operation presents a problem. In fact, theoretical analyses [191-193] show that in order to provide the continuous transfer of charge a diffusive conductor has to be much longer than the electron-phonon interaction length. For most materials at room temperature, this length is well above 10 nm (see, e.g., Ref. 64), i.e. much larger than the desirable size of the whole device. Moreover, the stray capacitance of such a resistor would be much larger than that of the island itself, reducing its single-electron charging energy and making room temperature operation impossible. As a matter of principle, electron hopping in quasi-insulators may ensure higher resistance at smaller resistor length; however, results of a recent analysis of quasi-continuous transport in this regime [194-196] also leads to pessimistic conclusions concerning the resistor size. For these reasons, prospects for room-temperature operation of any resistively-coupled single-electron devices do not look encouraging.

[7] It would be unfair to say that such poisoning is the specific problem of single-electronics. The electrostatic potential created by a single charged impurity in a typical dielectric is of the order of 1 Volt, the typical voltage scale for most electron devices including field-effect transistors. This is why this effect is important for *all* nanoscale devices. The only reason why it was noted in single-electron devices first is that these devices retain high charge sensitivity even if their islands are much larger (this is acceptable for low-temperature operation).

3.8. *Electrometry*

Now we are ready to discuss applications of single-electron devices. Unfortunately, due to space and time restrictions, I will not be able to discuss their use in fundamental physical experiments, and also in such interesting but narrow areas of electronics as single-electron spectroscopy (for a review, see, e.g., Ref. 144), dc current standards, and temperature standards [121, 122]. (For general reviews of analog and metrological applications of single-electron devices see, e.g., Refs. 131, 197 and 198.) I have to mention, however, one analog application which may become important for future integrated circuits, namely using single-electron transistors as ultrasensitive electrometers.

If the source-drain voltage V applied to a single-electron transistor is slightly above its Coulomb blockade threshold V_t, source-drain current I of the device is extremely sensitive to the gate voltage V_g. In fact, Figs. 9, 10 show that even the changes δV_g corresponding to sub-single-electron variations of the external charge lead to measurable variations of I. Calculations based on the orthodox theory have shown [199, 200] that the optimized charge sensitivity of such an electrometer is limited by its white (combined Johnson-Nyquist and shot) noise at the level

$$\delta Q \approx (k_B T R \Delta f)^{1/2} \times \begin{cases} 5.4C, & C_i << C, \\ 2.7C_i, & C_i >> C, \end{cases} \tag{25}$$

where $R = R_1 = R_2$ and $C = C_1 = C_2$ ($\ll e^2/k_B T$) are resistance and capacitance of each tunnel junction of the transistor, C_i is the effective output capacitance of the signal source (including the capacitance of the wires connecting it to the transistor), and Δf is the measurement bandwidth. This sensitivity is not at all impressive if C_i is large on the scale of C, which is typically very low (Fig. 13). On the other hand, if the source is so small and so close to the single-electron transistor that C_i of the order of C, the white noise limits the charge sensitivity only at an extremely low level of the order of $10^{-6} e/\sqrt{\text{Hz}}$. This is some 7 orders of magnitude better than sensitivity of the best commercially available instruments, and about 4

orders of magnitude more sensitive than specially designed low-temperature MOSFETs.

Tunnel barriers and electrostatic environment of single-electron devices always contain electron trapping centers and other two-level systems, each capable of producing "telegraph noise" – discrete, random low-frequency variations of the barrier conductance. An ensemble of these variations, with exponentially broad distribution of parameters, produces excess $1/f$-type noise; in single-electron transistors such noise may be very high, typically limiting the charge resolution at the level of the order of $10^{-4}e/\sqrt{\text{Hz}}$ for a-few-Hertz signal frequencies. The sensitivity may be reduced radically (to a level below $10^{-5}e/\sqrt{\text{Hz}}$ [201]) using special stacked geometry in which the single-electron island is lifted over the substrate that apparently hosts most of $1/f$ noise sources. Another way to reach a similar sensitivity is to modulate the transistor parameters, and pick out its output, at a GHz-range frequency [202, 203]. Such modulation cannot beat the white-noise-imposed limitation of sensitivity [204], but helps to avoid most of $1/f$ noise. In digital circuits it is easy to avoid the $1/f$noise by digital modulation, thus approaching the fundamental noise limit given by Eq. (25).

3.9. *Single- and few-electron memories*

The trade-off of advantages and drawbacks of single-electron devices is most favorable for memory applications (see, e.g., Ref. 205), because of the following reasons:

(i) for memories, the bit density is the most important single figure-of-merit,

(ii) low voltage gain of single-electron transistors may be tolerated,

(iii) simple rectangular-matrix architecture of the memory banks makes the exclusion of bad bits (say, with thresholds shifted by a single charged impurity) possible – see below.

The published suggestions for single-electron memory cells are based mainly on two approaches:

- using various modifications of the single-electron trap (Fig. 11) with either MOSFET or single-electron-transistor readout, and

- direct scaling down the cells of usual nonvolatile memories [206]. [8]

Physics of these two approaches is not much different: in both of them, insertion and extraction of a single electron to the trapping island ("floating gate"), i.e. write and erase operations, are achieved by the field suppression of the potential barrier separating the island from the electron source (word line). The only difference is that in the former case the barrier is created by the Coulomb repulsion of electrons in a short 1D tunnel junction array, while in the latter case the barrier physics is the usual conduction band offset a single thick tunnel barrier (e.g., ∼8 nm of SiO_2).

Numerous experiments with single-electron cells of both types were useful for the development of the field; in particular, room temperature operation of single cells has been demonstrated by several groups [26, 216-218]. However, because of the background charge randomness (see Sec. F above), these cells can hardly be sufficiently reproducible. In fact, a single charge impurity near the floating gate has an effect equivalent to an addition of (positive or negative) external charge $\Delta Q_0 \sim e$, and thus shifts the threshold for both write/erase and readout operations from their nominal values rather considerably. In the trap-type cells the same effect, in addition, may change the array proper-

[8] The latter approach includes an interesting proposal [207, 208] (see also Refs. 210-217) for relatively large (multi-electron) memory cells, using many (N) nanometer-scale silicon crystals rather than a single floating gate, in usual nonvolatile memories. The main advantage of this idea is that a single leaking defect in the tunneling barrier would not ground the whole stored charge, but only its minor ($\sim 1/N$) part. This may allow to use very thin tunnel barriers (which would be unreliable in the ordinary case) and thus to decrease the characteristic time of Fowler-Nordheim tunneling, which presents the lower bound for the write/erase cycle. A potential drawback of nanocrystalline floating gates is that the electric field of the surface of each crystal, which determines the Fowler-Nordheim tunneling rate, depends on the size and exact shape of the crystal and is basically unpredictable. This may provide an undesirable broadening of the statistics of the electric field at the surface of nanocrystal surfaces, and hence of write/erase thresholds, especially at any attempt to scale the cells below 10 nm, where the number N would be relatively small.

ties randomly [196]. [9]

Due to the regular structure of memory arrays, several ways of avoiding the random background charge effects are available. The idea suggested first [221] was to use the periodic character of the threshold characteristic of single-electron transistors for the cell contents readout. In this approach the memory structure may be very simple – see Fig. 15a. Binary 1 is stored in a relatively large floating gate in the form of a positive charge $Q = Ne$, with $N \sim 10$, while binary 0 is presented by a similar negative charge. (Since $N \neq 1$, the effect of random background charge on the floating gate is negligible.) Write/erase process is achieved by Fowler-Nordheim tunneling through a barrier separating the floating gate from the word line. Readout is destructive, and combined with WRITE 1 operation: if the cell contents was 0, during the WRITE 1 process the injected electrons ramp up the electric potential U of the floating gate, so that the external charge Q_0 of the readout single-electron transistor is ramped up by $N'e$, with $1 < N' < N$. Due to the fundamental periodicity of the transconductance (Fig. 10), this ramp-up causes N' oscillations of the transistor current. (If the initial charge of the floating gate corresponded to binary 1, the transistor output is virtually constant.) The current creates oscillations of voltage between two bit lines connected to SET source and drain. These oscillations are picked up, amplified, and rectified by an FET sense amplifier; the resulting signal serves as the output. The main idea behind this device is that the random background charge will cause only an unpredictable shift of the initial phase of the current oscillations, which does not affect the rectified signal.

This concept has been verified experimentally [222] using a low-temperature prototype of the memory cell. A very attractive feature of such SET/FET hybrid approach is a relatively mild minimum feature requirement: room temperature operation is possible with an electron addition energy of about 250 meV. Figure 13 shows that this level requires a minimum feature (SET island) size of about 3 nm, i.e. much larger than that required for purely single-electron digital circuits. The reason for this considerable relief is that in this hybrid memory the single-electron transistor works in the essentially analog mode, as a sense preamplifier/modulator, and can tolerate a substantial rate of thermally activated tunneling events. One drawback of this memory is the need for an FET sense amplifier/rectifier. However, estimates show that since its input signals have already been preamplified with the SET, one FET amplifier may serve several hundred memory cells within the ordinary NOR architecture [205, 206], and hence the associated chip real estate loss per bit is minor. The next drawback is more essential: the signal charge swing $\Delta Q_0 = N'e$ at the single-electron island should exceed e. Since the gate oxide should not bee too leaky to sustain acceptable retention time of the floating gate, its thickness should be at least a few nanometers. For silicon dioxide this gives a specific capacitance of approximately 1 μF/cm^2, corresponding to $\sim$10 nm^2 area per one electron (at 1 Volt). This prevents scaling of the island well below its maximum size ($\sim$3 nm).

Both these drawbacks may be avoided using a more complex memory cell (Fig. 15b, adapted from Ref. 223) in which the random background charge Q_0 of the island is compensated by weak capacitive coupling with the additional ("compensating") floating gate. The necessary few-electron charge of this gate may be inserted from an additional ("compensation") word line before the beginning of memory operation, when special peripheral CMOS circuit measures Q_0 of each single-electron transistor (this may be done in parallel for all bits on a given word line) and develops an adequate combination of bit line signals. During the actual memory operation, the compensating charge is constant, but may be adjusted periodically if necessary. Due to the compensation, the SET may be biased reliably at the steep part of its control characteristic (Fig. 10) and hence used for nondestructive readout of the cell state. Moreover, the signal charge swing may be relatively small, $\Delta Q_0 \sim (k_B T/E_a)e$, conveniently decreasing with the island size. This may allow the single-electron

[9] Memory cells based on single- (or few-) electron trapping in grains of nanocrystalline MOSFET channels [219, 220] have even larger threshold spreads due to random locations of the grains and random transparency of tunneling barriers.

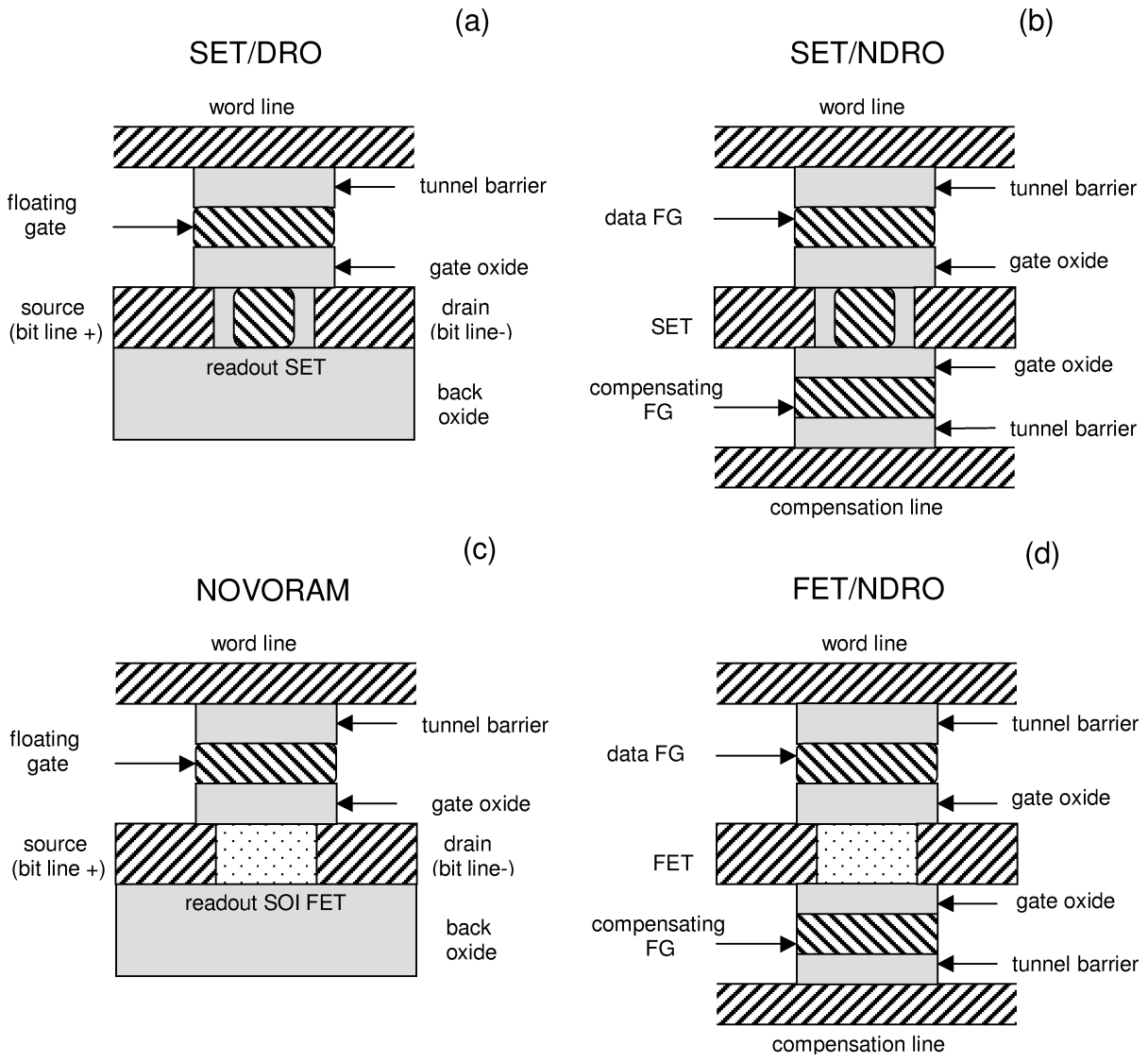

Fig. 15. Suggested few-electron memory cells (schematically): (a) cell with destructive SET readout (SET/DRO, [221]); (b) SET-readout cell with background charge compensation, making non-destructive readout possible (SET/NDRO, [223]), (c) cell with nondestructive FET readout which may also be used in NOVORAM [226], and (d) a similar cell with compensation of FET parameter variations [99].

transistor island to be scaled down to ∼1 nm, with the ∼3 nm floating gate storing just 2 to 3 electrons.

An alternative approach to few-electron memories is to scale down the generic structure of nonvolatile memory cells [205, 206] using a nanoscale MOSFET for readout – see Fig. 15c. As was already mentioned in Sec. 2, advanced MOSFETs with ultrathin silicon-on-insulator channel may be scaled down to ∼10 nm sustaining high performance and reproducibility. Further scaling down would lead to large random fluctuations of the threshold voltage, that may be again corrected using a back floating gate (Fig. 15d).

Estimates of the maximum density of the SET/FET hybrid memories may be carried out under the assumption that the SET islands are fabricated by some self-aligned method, so their size is independent of the wiring line half-pitch F. In this case the area of the cells shown in Fig. 15 is essentially independent of the island size and ranges between $6F^2$ and $8F^2$, where F is the minimum feature size of a given technology. The cell type determines the possible limits of this scaling: for the SET/DRO (Fig. 15a) it is confined in a narrow range around 3 nm, while for the SET/NDRO (Fig. 15b) F may be scaled down to ∼1.5 nm (line patterning permitting). For the FET/NDRO cells shown in Fig. 15c,d, the range of F does not have an upper bound; for the simple cell shown in Fig. 15c the lower bound for F is about 5 nm, while the more complex cell with background charge compensation (Fig. 15d) can be scaled down to $F \sim 2$ nm, i.e. to the cell area about 30 nm^2. This should allow memories with density well beyond 10^{12} bits/cm^2, enabling chips with multi-terabit integration scale. These exciting prospects are, however, contingent on the development of nm-scale fabrication technologies.

Another problem with all the memory cells described above is the slowness of the write/erase operations. In fact, they rely on the process of Fowler-Nordheim tunneling, similar to that used in the ordinary nonvolatile, floating-gate memories [206]. The standard 8-to-10-nm SiO_2 barriers used in such memories do not change transparency fast enough to allow floating gate recharging than in ∼1 μs even if the applied electric field is as high as

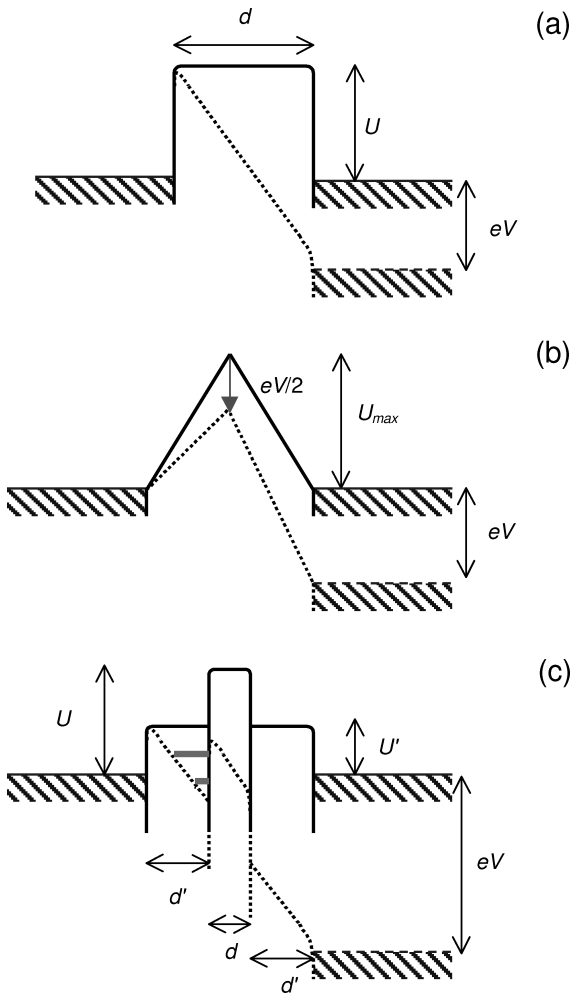

Fig. 16. Tunnel barriers options: (a) usual, uniform barrier; (b) ideal, triangular crested barrier; and (c) trilayer crested barrier [224]. Thick horizontal lines show (schematically) the subbands formed at large values of applied voltage, and enabling resonant tunneling through the barrier.

∼10 MV/cm, close to the breakdown threshold. (A reduction of the barrier thickness or height makes retention time too short.) Such a long write/erase cycle is acceptable for typical applications of flash memories [206], but is too long to replace DRAM in bit-addressable memories.

This problem can be solved using special layered ("crested") barriers [224] (Fig. 16). Calculations show [224-226] that trilayer crested barriers may combine a 1-ns-scale write time with a ∼10-year retention time, at apparently acceptable electric

fields (about 10 MV/cm) – see Fig. 17. This gives hope that such barriers may have extremely high endurance necessary for RAM applications. Moreover, ratio of the necessary write voltage (V_W) to the highest retention voltage (V_R) for such barriers may be well below 3. This condition ensures that the disturb effects on self-selected cells are small enough even in the simplest memory architectures [226].

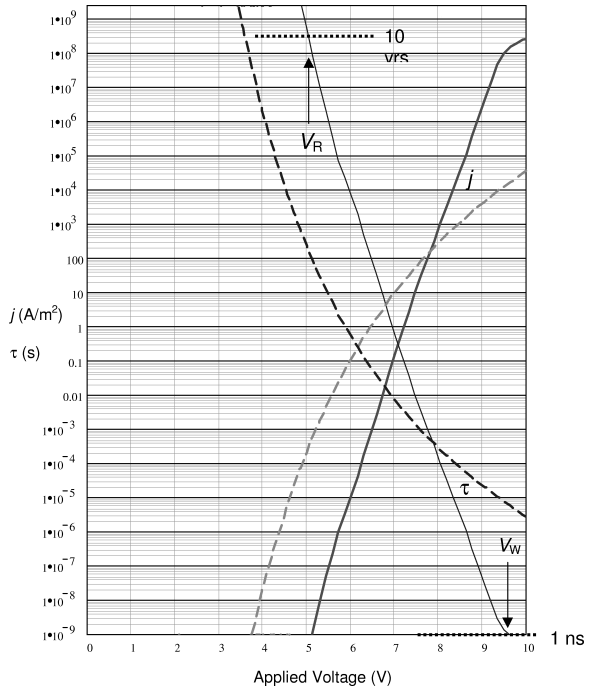

Fig. 17. Calculated current density j and floating gate recharging time constant τ as functions of the applied voltage V for a uniform tunnel barrier (10 nm of SiO_2, dashed lines) and a trilayer crested barrier (5 nm of thermally-grown and post-annealed Al_2O_3 sandwiched between two plasma-grown 2.5-nm Al layers, solid lines) [99].

Thus, if CMOS-compatible crested barriers are implemented, the few-electron cells shown in Fig. 15 may be used as random-access memories, replacing the generically unscalable DRAM. Moreover, crested barriers may enable the so-called Nonvolatile Random Access Memory (NOVORAM [224-226]) with a very simple cell structure (Fig. 15c) and architecture (Fig. 18), which may be able to compete with DRAM and possibly SRAM even at the current technology level (minimum feature size F of the order of 100 nm), and be scaleable all the way down to $F \sim 3$ nm. Some further decrease of F is possible by dropping the nonvolatility requirement and organizing cell refresh with period ~ 1 s, because in this case the crested barrier and gate oxide may be thinned by $\sim 50\%$.

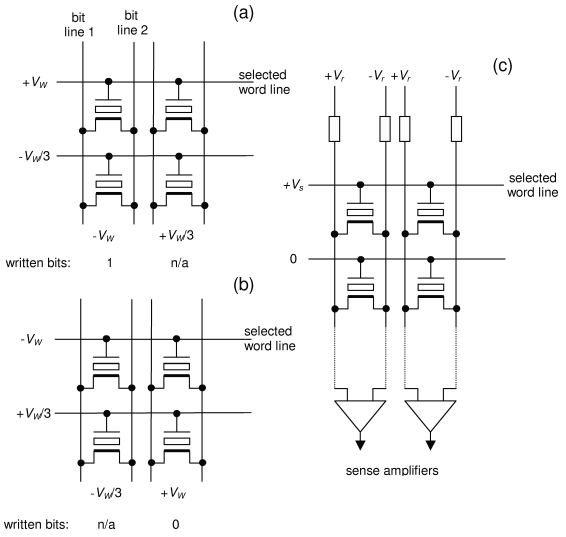

Fig. 18. Possible architecture of NOVORAM and applied voltages necessary for: (a) WRITE 1, (b) WRITE 0, and (c) READ operations [99].

An alternative way to speed up the write/erase operation in floating gate memories is to replace the tunnel barrier with a transistor. Such two-transistor (2-T) memory cells were suggested long ago [227]; for a while they could not compete with DRAM because of the much larger cell size. The apparent impossibility to scale DRAM much below 100 nm has led to a recent revival of R&D work in this direction, and several new versions of the 2-T memories have been suggested (and some explored experimentally):

1. Single-electron trap with an additional gate controlling the potential profile high of the array, and readout using either a single-electron-transistor [228], or a nanoscale MOSFET [229-232]. Unfortunately, neither approach addresses the problem of random background charge. (Single-electron arrays, in addition, are themselves sensitive to this effect [196, 233].) Notice that in experiments with semiconductor single-

electron arrays the Coulomb blockade mechanism of electron transport control coexists (and was repeatedly confused) with the usual field-effect mechanism. In this case we are speaking about the effective replacement of the single-electron array by an MOSFET with very thin (and hence not very uniform) channel.

2. The authors of a recent work [234] have made this replacement conscientiously, arguing that their 2-nm thick silicon channel should have lower parasitic source-to-drain leakage, because quantum confinement increased the effective bandgap. Indeed, their MOSFETs exhibited leakage current as low as 10^{-19} A, sufficient to keep the cell retention time above 100 ms, i.e., the typical DRAM refresh time. A memory similar in structure, but with a special "stacked" vertical FET [235] for fast write/erase is called Phase-State Low Electron Drive Memory, or PLEDM [236] – see also [237]. The channel of such a transistor incorporates three horizontal silicon nitride tunnel barriers partitioning the channel into several parts connected in series. Such separation improves transport control of the channel potential (and hence of the electron transport) by the surrounding gate; the considerable loss of ON current caused by these barriers is tolerable because the MOSFET should only recharge a very small capacitance of the charging node. Unfortunately, in such vertical MOSFETs, with their relatively large channel cross-section, getting acceptably small leakage current (and hence retention time) may be a difficult problem.

Though the approach [234] looks very interesting, I believe that the successful implementation of silicon-compatible crested barriers [224] will make NOVORAM, with its simple structure and small cell area, a more promising option.

The single- and few-electron memories will have to compete not only with each other, but with several other prospective memory concepts.

In *ferroelectric memories*, information is stored as a sign of electric polarization of a layer of a ferroelectric material. This polarization may be read out either destructively (as in DRAM) or non-destructively (e.g., if it controls a readout FET) [238]. Strong features of these memories include a simple cell structure (and, as a result, small cell area), potential nonvolatility, and fast write/erase time (some materials have internal re-polarization time well below 1 ns). In terms of immediate practical introduction, the complexity of ferroelectric materials and their compatibility with the generic CMOS process is the main challenge. However, in the long term, scalability of these memories may be a larger problem. With the decrease in area, the height of the energy barrier $\Delta U \approx P_s E_c V$ separating two polarization states (where P_s is the saturation polarization, E_c the coercive electric field, and V the ferroelectric layer volume) decreases and should finally become comparable with the thermal fluctuation scale $k_B T$, resulting in random cell switching. For typical ferroelectric film parameters $P_s \sim 50$ μC/cm^2 and $E_c \sim 5$ V/μm [206], such spontaneous switching should become rather noticeable ($\Delta U \sim 300\ k_B T$) at $V \sim 20{\times}20{\times}2$ nm^3, even if P_s and E_c do not degrade with size. Thus, if no new breakthroughs are made, ferroelectric memory cells can hardly compete for terabit applications, which is the main promise of single- and few-electron memories.

In *magnetic memories* [238, 240], bits are stored in the form of thin film magnetization. These memories share almost all the advantages and drawbacks of ferroelectric cells listed above (low 0/1 output signal value is an additional issue). However, because of the essentially similar physics, scaling of such cells to nanoscale may again be a problem. In this case, $\Delta U \sim B_s H_c V$ (where $B_s \sim 2$ T is the saturation magnetization and $H_c \sim 10$ Oe the coercive magnetic field) drops below 300 $k_B T$ at even larger volume ($\sim 30{\times}30{\times}10$ nm^3).

Memory cells based on *structural phase transition* include notably the "Ovonics Unified Memory" (*OUM*) [241, 242]. In an OUM cell, a chalcogenide alloy (GeSbTe) is switched from a conductive crystalline phase to a highly-resistive amorphous phase under the effect of heating by current passed through a special heater. Though the chalcogenide materials are relatively complex, considerable progress in their deposition has been made in the course of development of CD-RW and DVD-RW technologies. As a result, OUM cells with surprisingly high endurance (up to 10^{13} cycles) have been demonstrated. OUM problems include relatively long write/erase time (reportedly, close to 100 ns, i.e., considerably longer than that

of DRAM). Unfortunately, I am not aware of any published experimental data or reliable theoretical results sufficient to evaluate the dependence of the retention time on the storage region volume, and thus evaluate prospects of OUM scaling into the terabit range.

Finally, *single-molecular memories* may be based on various background physics. For example, they may be just molecular implementation of single-electron memories discussed above (Fig. 15). However, there is an alternative possibility: to employ molecular conformation changes [159, 163, 315]. In this case the molecule has internal bistability that manifests itself as a hysteretic region on the molecule as a two-terminal device. Very recently, this approach allowed the first demonstration of an 8×8 memory matrix, where each bit was stored in a state of many similar molecules, sandwiched in parallel between two crossing 40-nm metallic wires [245]. Since the technical details of this work have not yet been published, it is too early to evaluate the possible speed, retention time, and reliability of such memory cells. However, I will discuss this general approach more in Sec. 4 below.

3.10. *Electrostatic data storage*

A combination of single-electron transistors with crested barriers may be used not only in terabit-scale memories, but also in ultra-dense electrostatic data storage systems [224] – see Fig. 19. In this "ESTOR" design, a single-electron transistor, followed by a MOSFET amplifier, a few microns apart, would be fabricated on a tip-shaped chip playing the role of a READ/WRITE head. The data bits are stored as few-electron charge trapped in a group of nanoscale conducting grains deposited on top of a crested tunnel barrier. It is important that since each bit is stored in a few (∼10) grains, their exact shape and location are not important, so the storage medium production does not require any nanofabrication.

WRITE is performed by the application of the same voltage V_W to both input terminals, relative to the conducting ground layer of the moving substrate. The resulting electric field of the tip induces rapid tunneling of electrons from the ground through the crested barrier into a ∼30-nm-wide group of grains. For READ, the single-electron transistor is activated by source-drain voltage $2V_R \geq V_t$. In this state it is very sensitive to the electric field created by the group of charged grains it is being flown above.

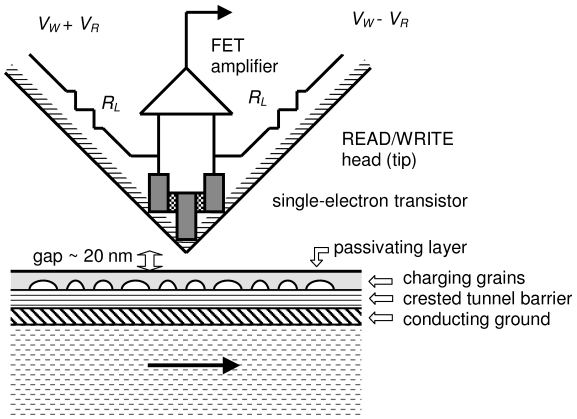

Fig. 19. Proposed electrostatic data storage system ("ESTOR") with hybrid SET/FET readout [224].

Simple estimates show that with a 20-nm tip-to-substrate distance (close to those already implemented in the best present-day magnetic storage systems), the electrostatic system is capable of a density ∼3 Terabits per square inch, i.e. at least an order of magnitude than the best prospects for the magnetic competition of which I am aware. The use of crested barriers may provide a very broad bandwidth of both WRITE and READ operations, up to 1 Gbps per channel, possibly quite adequate even for this enormous bit density.

Notice that for this particular application the difficulties of fabrication of room-temperature single-electron transistors, outlined above, are not a major concern, because one would need just one (or a few) transistors per system, and slow fabrication techniques (like the scanning probe oxidation) may be acceptable. Moreover, like in single-electron memories, the transistor would work as an analog amplifier, so that the single-electron island size of ∼3 nm would be sufficient.

Recent experiments [246] may be considered as the first step toward the implementation of this idea.

3.11. *Logic circuits*

Most suggestions of logic circuits based on single-electron devices may be referred to one of two groups.

1). In circuits of the first, *"voltage state"* group, single-electron transistors are used in CMOS-like circuits. This means that the single-electron charging effects are confined to the interior of the transistor, while externally it looks like the usual electronic device switching persisting currents, with binary unity/zero presented with high/low dc voltage levels (physically not quantized). This concept simplifies the circuit design which may ignore all the single-electron physics particulars, except the specific dependence of the drain current I on the drain-to-source voltage V and gate-to-source voltage V_g - see Figs. 9 and 10.

Analyses of this opportunity has shown that due to the specific shape of this dependence (oscillating transconductance), both resistively-coupled [124] and capacitively-coupled [247] single-electron transistors allow a very simple implementation of CMOS-type inverters, without a need for two types of transistors (like n-MOSFETs and p-MOSFETs in the standard CMOS technology). On the other hand, peculiarities of functions $I(V, V_g)$ makes the exact copying of CMOS circuits impossible, and in order to get substantial parameter margins, even simple logic gates have to be re-designed. Such circuits (Fig. 20) may operate well within a relatively wide window of parameters R, C, and V_{DD} [248, 249]. Even after such optimization, the range of their operation with acceptable bit error rate shrinking under the effect of thermal fluctuations as soon as their scale $k_BT$reaches approximately $0.01E_a$. (For other suggested versions of the voltage state logic [247, 250], the temperature range apparently is even narrower). The maximum temperature may be somewhat increased by replacing the usual single-island single-electron transistors with more complex devices with 1D-array "channels" and distributed gate capacitances [251]. However, this would increase the total transistor area, at the given minimum feature size.

A disadvantage of voltage state circuits is that neither of the transistors in each complementary

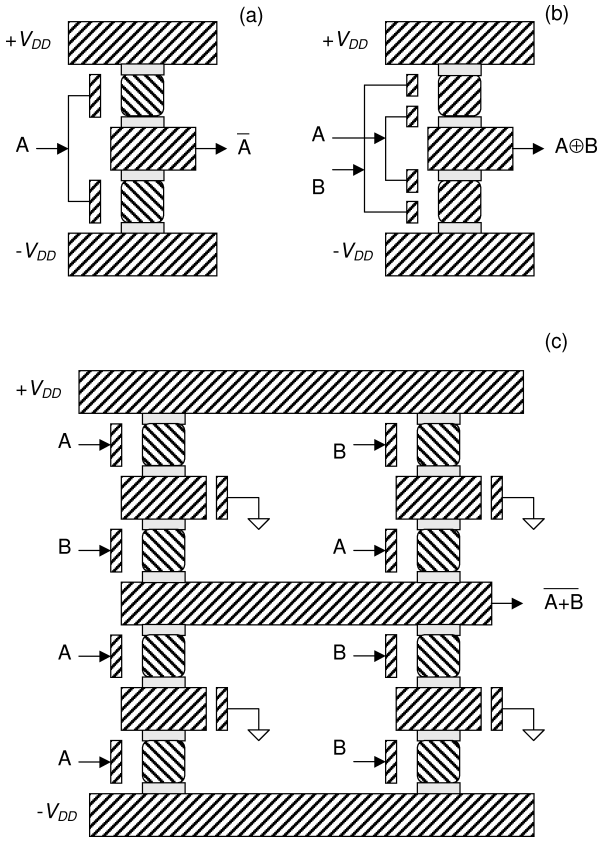

Fig. 20. Basic gates of the complementary SET logic family using capacitively-coupled single-electron transistors [249]: (a) inverter; (b) XOR, and (c) NOR/NAND.

pair is closed too well, so that the static leakage current in these circuits is fairly substantial, of the order of $10^{-4}e/RC$ [248]. The corresponding static power consumption is negligible for relatively large devices operating at helium temperatures. However, at the prospective room-temperature operation this power becomes on the order of 10^{-7} Watt per transistor. Though apparently low, for the hypothetical circuits which would be dense enough ($>10^{11}$ transistors per cm^2) to challenge the prospective CMOS technology, this number gives an unacceptable static power dissipation density (>10 kW/cm^2).

2) The power dissipation problem may be avoided, to a large extent, by using *"charge state"* logic circuits in which single bits of information are presented by the presence/absence of single

electrons at certain conducting islands throughout the whole circuit. In this case the static currents and power virtually vanish. This approach has been explored theoretically since 1987 [252], and several few families of charge state logic circuits have been suggested and analyzed [139, 140, 253-283]. In most suggestions, an electron is confined in a cell consisting of one or a few islands, while the logic switching is achieved via electrostatic (or spin [259, 264, 271]) coupling of the cells. Another classification of single electron logics may be based on where they take the energy necessary for logic operations: from dc power supply [252, 261, 269], ac power supply (also playing the role of global clock) [138, 140, 260, 262, 263, 274, 278-283], or just from the energy of an external signal [254-259, 265, 267, 271, 272]. [10]

Only a few of these concepts have been analyzed in detail, especially at finite temperatures. To my knowledge, the most robust charge-state logic circuits suggested till now are those based on the single-electron parametrons (see Sec. 3.5 above). Figure 21 shows a possible shift register based on the parametrons [139]. The direction of the shift of the central island of each next device is shifted by $\pi/3$ within the yz plane. The circuit is driven by electric field $\boldsymbol{E}_c(t)$ rotating in the same plane and providing the periodic switching on the SET parametrons, with an appropriate phase shift. As a result, each digital bit (one per three cells) is being shifted by 3 cells along the structure each clock period. Majority logic gates, sufficient for arbitrary logic circuits, may be implemented in the same way [140]. Geometric modeling and numerical simulation of these circuits within the framework of the orthodox theory have shown that they may operate correctly within approximately $\pm 20\%$ deviations from the optimal clock amplitude. Estimates show that the maximum operation temperature of these logic circuits is of the order of that of voltage mode circuits, i.e. of the order of $0.01E_c/k_B$, if the bit error rate is in the practically acceptable range (below $\sim 10^{-20}$).

A new, potentially useful feature of the charge state logics is the natural internal memory of their "logic gates" (more proper terms are "finite-state cells" or "timed gates"), thus combining the functions of the combinational logic gates and latches. This feature makes natural the implementation of deeply pipelined ("systolic") and cellular automata architectures. The back side of this advantage is the lack of an effective means of transferring a signal over large distances: crudely speaking, this technology does not allow passive wires, just shift registers.

3) Within the framework of our classification, the so-called *phase-mode logic* [284-287] should be placed in a separate category, because in these circuits the information-keeping cell are coupled by rf signals, carrying the binary information coded by either of two possible values of the rf signal phase. The elementary cell of this logic is essentially a relaxation oscillator generating SET oscillations with frequency $f_{SET} = I/e$ [123, 286], phase locked by an external reference with frequency $2f_{SET}$. Such "subharmonic" phase locking allows two possible phases of oscillations, which differ by π; the oscillation state with each phase has exactly the same amplitude and is locally stable, i.e. can be used to code a single bit of information. Moreover, the oscillator can impose its phase upon a similar, adjacent oscillator that had been turned off temporarily (e.g., by turning off its power supply current I), and is now being turned on. Thus information may be transferred along a row of oscillators, working as a shift register, very much similar to that consisting of single-electron parametrons – see Fig. 21. Majority gates can be organized similarly, enabling the implementation of arbitrary logic functions.

A theoretical advantage of this logic family is

[10] The last category includes the so-called Quantum (or Quantum-dot) Cellular Automata (QCA) based on ground state computing. This concept had been controversial from the very beginning [260, 270-272], because it leads to the problem of system trapping for exponentially large (read infinite) time in intermediate metastable states – for a detailed discussion, see the recent publications [282, 283]. Eventually the concept was repudiated by its authors – see the introduction section in Ref. 141. The second variety of logics based on similar cells ("adiabatic" or "clocked" QCA) [263] is very close to single-electron parametron circuits (that had been suggested earlier [139]), besides that the QCA cells have a more complex structure and as a result lower speed and narrower parameter margins [282, 283].

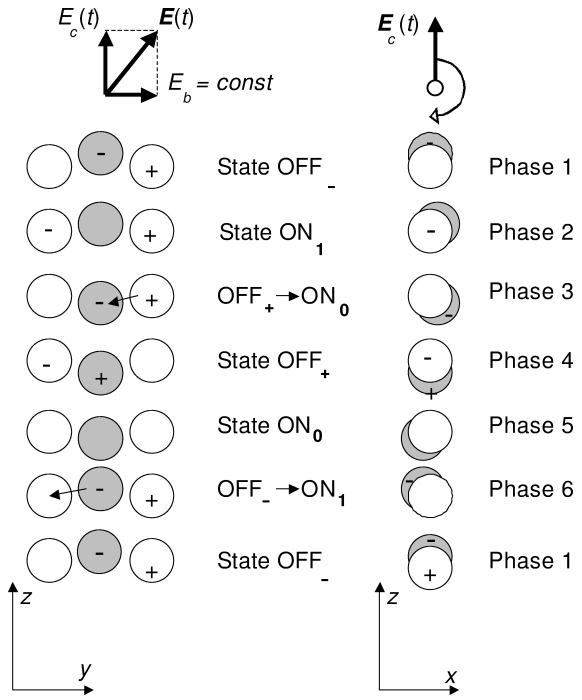

Fig. 21. A shift register based on single-electron parametrons.

the potential insensitivity of its cells to the random background charges, if a resistor with quasi-continuous charge transfer is used for the insertion of dc current I. Unfortunately, as was discussed in Sec. 3.7 above, the known implementations lead to very large size of such resistors, making the implementation of practical (room-temperature) phase-mode logic circuits hardly feasible.

Unfortunately, *all* the single-electron logic circuits discussed above face two crucial problems:

(i) In order to ensure an acceptable bit error rate, they require single-electron islands with very high electron addition energy, of the order of 100 k_BT. For room temperature operation, this means the island size below 1 nm (Fig. 13). At the same time, parameter margins of these devices are not very high (∼20% or lower). This means that the critical dimension control (on 3σ level) should be of the order of 0.1 nm , i.e., even more tough than for nanoFETs.

(ii) Traditional logic circuit architectures (e.g., microprocessors, digital signal processors, etc.) do not allow a straightforward determination and exclusion of faulty logic gates. This is why the effect of random background charge (Sec. 3.7 above) cannot be circumvented as simply as in memories (Sec. 3.9).

4. CMOL: Devices, circuits and architectures

4.1. *The concept*

The hard challenges faced by both sub-10-nm field-effect transistors and room-temperature single-electron logic circuits make me believe that the only chance of breakthrough beyond the 10-nm frontier is the complete change of the integrated circuit paradigm. The approach I am advocating is the transfer to "CMOL" hybrid circuits combining CMOS components and molecular (e.g., single-electron) devices, interconnected by nanowires (Fig. 22) [292]. [11]

The lower level is occupied by a CMOS stack. The transistor density of this sub-system cells may be relatively low, of the order of 10^8 cm^2. (The implementation of this density would require the 45-nm-node technology, which should be commercially available by the end of this decade [1].) Vertical plugs connect the CMOS circuit to I/O pins (including those providing power supply for the whole circuit), CMOS wiring, and the next circuit level, occupied by nanowiring.

In order to sustain the necessary density of molecular devices, these wires should be extremely narrow (a few-nm half-pitch). Simultaneously, the wire resistance per unit length should be not too high, below ∼10 MΩ/μm, to sustain relatively high operation speed. Because of this, as well as the necessary chemical compatibility with molecular devices of the top level, I prefer to think about explicitly patterned gold wires rather than molecular-wire options. Of course, patterning with this resolution is very challenging; however, a CMOL circuit would use just a few (say, two) layers of uniform, parallel, nanowire sets in two

[11] The hybrid approach makes CMOL rather different from several earlier concepts of molecular electronics circuits - see, e.g., Ref. 289.

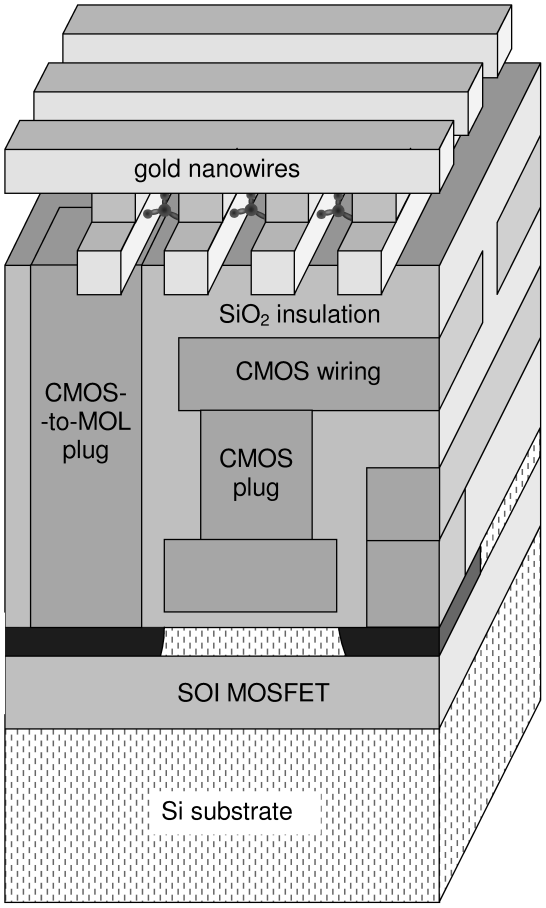

Fig. 22. General concept of CMOL (CMOS/nanowire/MOLecular hybrid) circuit.

mutually perpendicular directions. Such a simple line pattern does not require nanometer-scale level alignment, and may be formed, e.g., by nanoimprinting [290]. (See also Chapter 3 of this collection.)

Finally, the nanowire layers are connected by self-assembling molecular devices having the smallest individual footprint and hence, the highest possibly density. With the 3-nm nanowire half-pitch (close to the limit imposed by wire-to-wire tunneling, see Sec. 2.4 above), the density of active molecular components would exceed 3×10^{12} functions per cm^2. (Just for comparison, Encyclopedia Britannica contains about 10^{10} bits of information.)

I believe that the advantage of the CMOL approach is that if allows an optimum combination of strengths of its components: robust and universal CMOS circuits may take care of the functions requiring high reliability, high voltage gain and high ON currents, in particular signaling over long (1-cm-scale) distances, as well as I/O functions. On the other hand the molecular devices will sustain simple functions requiring highest integration scale. But most importantly, self-assembly of molecular devices may allow to keep the total fabrication costs of CMOL chip comparable with today's industrial level, thus avoiding the largest threat to the Moore-Law progress.

4.2. *Defect-tolerant architectures*

Even if/when VLSI CMOL circuits are successfully implemented, the yield of self-assembled molecular devices will hardly approach 100%. Moreover, as was discussed in previous sections, a-few-nm-scale devices will always be sensitive to random charged impurities, regardless whether they use FET-like or SET-like physics for electron transport control. Hence the CMOL approach requires defect-tolerant architectures that would allow to either tolerate or exclude bad devices. The most evident opportunity here is to use the molecular level as massive embedded memory for CMOS-based circuits [291, 308]. (Several industrial groups are already working in this direction – see, e.g., Ref. 245.) For more advanced circuits, there are several opportunities, [12] and I will describe just one of them, which is the focus of the recent work of our Stony Brook group [292-294, 316].

It is well known that neuromorphic networks [295-302] that mimic the basic functions of the cerebral cortex [303, 304] may be used for extremely efficient information processing. For example, the human brain can carry out an almost perfect recognition of a visual image in approximately 100 milliseconds, i.e., in just ~30 elementary "ticks" (neural spike durations) of the cortical circuitry. For comparison, the best modern microprocessor, using the best software available,

[12] For example, an interesting idea of defect-tolerant digital computing based on look-up-tables (essentially, large memories) is discussed in Ref. 310.

can perform a less reliable recognition on the scale of minutes, i.e. in $\sim 10^{12}$ of its 1-ns-scale "ticks" (clock periods).

The implementation of this advantage in artificial systems may require, however, VLSI circuits with the number of active components comparable with that in the cerebral cortex. The number of neural cells in mammal's cortex ($\sim 10^{10}$ [303, 304]) is not overwhelming, but they feature a very large *connectivity*: each cell is directly connected, on the average, to $M \sim 10^4$ other cells. Each connection is served by an active device, the synapse. Thus the number of active functions in the cerebral cortex is of the order of 10^{14}, and only molecular nanoelectronics gives the first hope for the fabrication of comparable structures on acceptable scale of circuit size, power dissipation and (yes, Virginia!) fabrication costs. Within the CMOL approach (Fig. 22), CMOS may take care of functions of neural cell bodies ("somas"), but synapses should be implemented using more dense molecular devices.

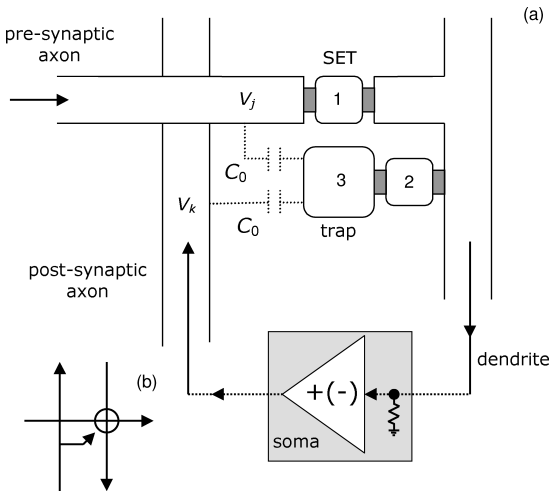

Fig. 23. (a) The simplest three-terminal single-electron latching switch and (b) its circuit notation.

Figure 23a shows the three-terminal version [294] of earlier suggested [292] device that is a simple combination of the single-electron transistor and the trap, working together as a "latching

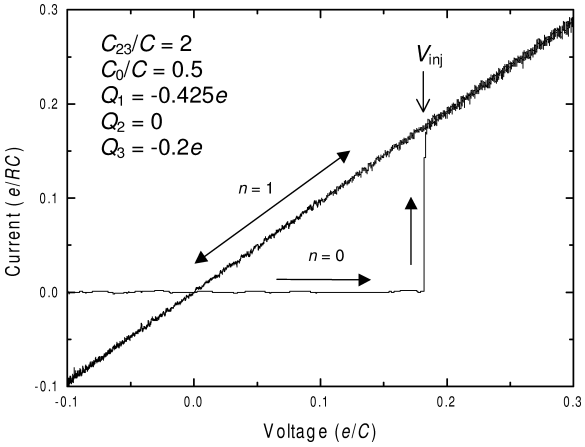

Fig. 24. DC $I - V$ curve of the latching switch, calculated using the "orthodox" theory [292]. For the three-terminal switch (Fig. 23), the effective applied voltage is $V_j + V_k$.

switch".[13] The device consists of three small islands connected by four tunnel junctions. Island 1, together with input and output wires serving as source and drain, forms a single-electron transistor. Islands 2 and 3 form a single-electron trap (cf. Fig. 11a), with trapping island 3 capacitively coupled to the SET island 1, thus playing the role of a single-electron floating gate. If the effective voltage $V \equiv V_j + V_k$ applied to the device is low, the trap in equilibrium has no extra electrons and its total electric charge is zero. As a result, the transistor remains in the Coulomb blockade state, and input and output wires are essentially disconnected. If V is increased beyond a certain threshold V_{inj} (which should be lower than the Coulomb blockade threshold voltage V_t of the transistor), one electron is injected into the trap. In this charge state the Coulomb blockade in the transistor is lifted, keeping the wires connected at any V. However, if the node activity (voltage V) is low for a long time, either thermal fluctuations or

[13] Notice also that island 2 is not really necessary and may be replaced by a thicker tunnel barrier – see the discussion in the end of Section 3.4. The device so modified is essentially a three-terminal version of a four-terminal device that had been discussed qualitatively in Ref. 310. Earlier suggestions to use single-electron devices in neuromorphic networks [305, 309-311] focused on the implementations of the somatic functions that should be, in my opinion, left for more robust CMOS circuits.

co-tunneling eventually kick the trapped electron out of the trap and the transistor closes, disconnecting the wires. Figure 24 shows typical results of numerical simulation of the latching switch dynamics with a perfect background charge of the SET island (it may be adjusted by voltage applied to an additional global gate). [14]

Figure 25 shows the possible molecular implementation of the three-terminal latching switch [294]. The role of single-electron island is played by diimide groups well known for their acceptor properties (see, e.g., Ref. 306); OPE bridges [307] are used as tunnel junctions, while thiol groups play the role of "alligator clips" that should allow the molecule to self-assemble from solution on gold nanowires [308].

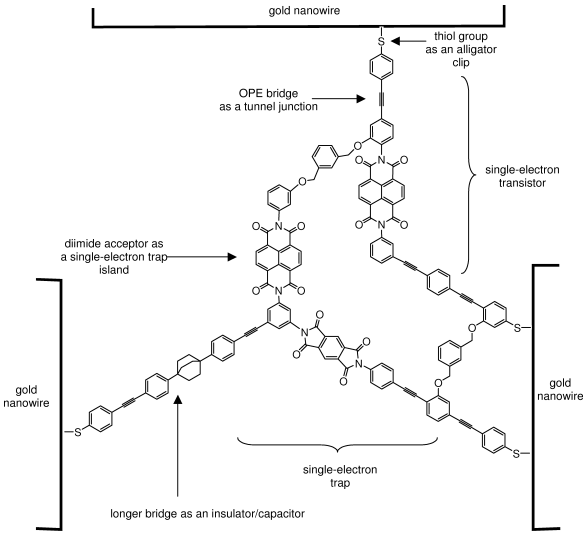

Fig. 25. Possible molecular implementation of the latching switch shown in Fig. 23 [294].

We have shown [294] that a group of four devices shown in Fig. 23a may implement a "Hebbian" synaptic function: the net synaptic weight is strengthened when the post-synaptic activity immediately follows the pre-synaptic activity [295-302].

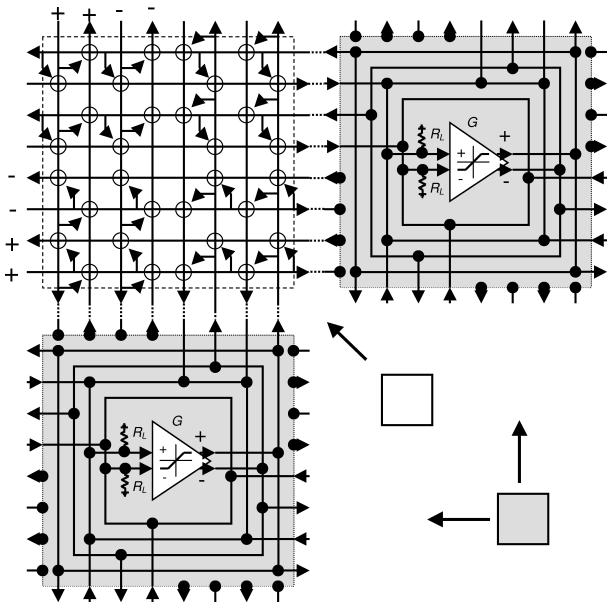

Fig. 26. Structure of a distributed crossbar network (CrossNet), with the synaptic and somatic (gray) plaquettes.

Figures 26 and 27 show the general structure, and two most promising species of the so-called distributed crossbar arrays for neuromorphic networks ("CrossNets") based on such synapses [293]. In each CrossNet, somatic cell interfaces (gray cells) are embedded sparsely into a 2D array of synaptic plaquettes, each containing 8 synapses (each synapse is a group of 4 latching switches). [15] As a result, each somatic cell is hard-wired to a large number, $4M$, of other somas, with the binary synaptic weights controlling which of these connections are currently active. Vice versa, the signal activity of the network determines whether the synapses are open or closed, though the state of any particular synapse is also affected by the underlying randomness of single-electron tunneling.

In InBar (which currently looks like the most promising CrossNet option), the gray cells sit on a square lattice inclined (hence the name) relative

[14] These simulations were based on the "orthodox" theory of single-electron tunneling [115]. Though this theory gives a good qualitative guide to properties of even molecular-size single-electron devices, their serious quantitative theory should be based on *ab initio* calculations of molecular orbitals – see, e.g., Refs. [181-186] and Chapter 6 of this collection and the transport theory taking into account the electron state discreteness [142-144].

[15] The CMOS-implemented somatic cell itself may be much larger than the synaptic plaquette, occupying all the chip area between two adjacent gray cells.

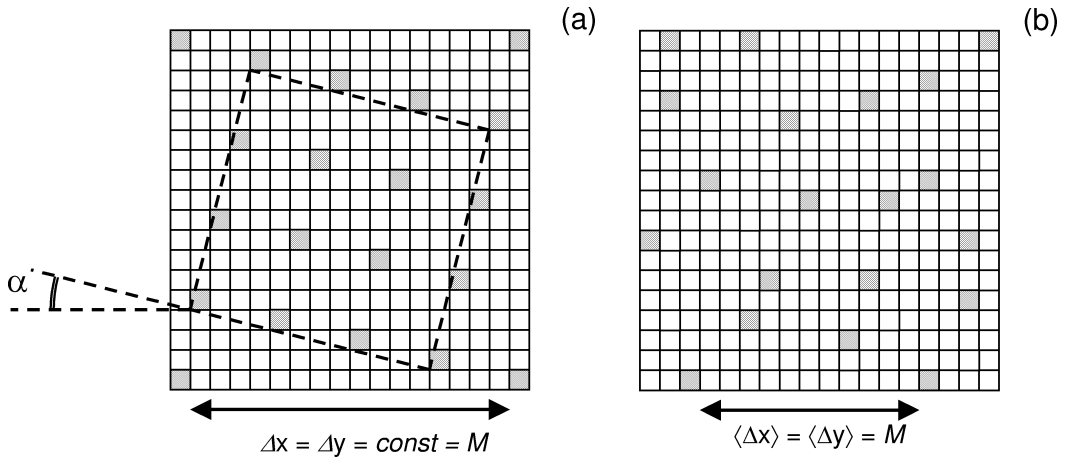

Fig. 27. Two CrossNet architectures: (a) InBar and (b) RandBar. M is the connectivity parameter. For the plaquette notation, see Fig. 26.

to the synaptic plaquette array. The incline angle α determines the network connectivity parameter: $M = 1/\tan^2\alpha$. As can be seen from Figs. 26 and 27, the total "Manhattan" (dendritic plus axonic) distance between each pair of coupled cells in InBar, measured in synaptic plaquettes, is the same and equals to M. On the other hand, in RandBar the gray cell terminals are distributed randomly. This creates the Poissonian distribution of intercell distances, so that there is a small amount of very long interconnects. We see that this property creates some difference in network dynamics, but are still not sure whether this difference is substantial for information processing.

Our collaboration is currently working on the molecular implementation of CrossNet synapses and, in parallel, on numerical simulation of limited fragments of these networks on usual supercomputers, trying to train them to perform various functions. The important rule of this game is that, like in the future hardware implementations, the external tutor system has access only to (sparse) somatic cells, rather that to individual synaptic weights. This restriction, as well as deeply-recurrent nature of CrossNets and the statistical character of single-electron synaptic weights, does not allow for the straightforward use of the well-known methods of neuromorphic network training [295-302]. However, we have already succeeded to demonstrate [316] that in spite of these restrictions, as well as quasi-local coupling of somatic cells (finite M), CrossNets with Hebbian synapses may operate rather well as Hopfield networks. (Figure 28 shows the example of black-and-white image recognition in this mode.)

For more advanced applications, such as pattern classification and feature detection, the continuous-mode training is necessary. Very recently, we suggested [316] a way for such training, using chaotic self-excitation of CrossNets at large values of somatic amplifier gain. If these expectations are confirmed, CrossNets may work as pattern classifiers with very impressive characteristics. Estimates show [292] that at the neural cell connectivity $4M = 10^4$ (comparable with that of the human cerebral cortex) and a nanowiring half-pitch of 3 nm, the neural cell density may be as high as 10^7 cm^{-2}. The estimated time of signal propagation between the neural cells of the order of 20 ns, at high but acceptable power dissipation (100W/cm^2). This speed is approximately 6 orders of magnitude higher than that in biological neural networks. It is expected that the introduction of such circuits will create a new electronic market comparable, if not larger than the PC market.

This success would pave the way toward much more ambitious goals. It seems completely plausible that a cerebral-cortex-scale CrossNet-based system (with $\sim 10^{10}$ neurons and $\sim 10^{14}$ synapses, that would require $\sim 30\times 30$ cm^2 silicon sub-

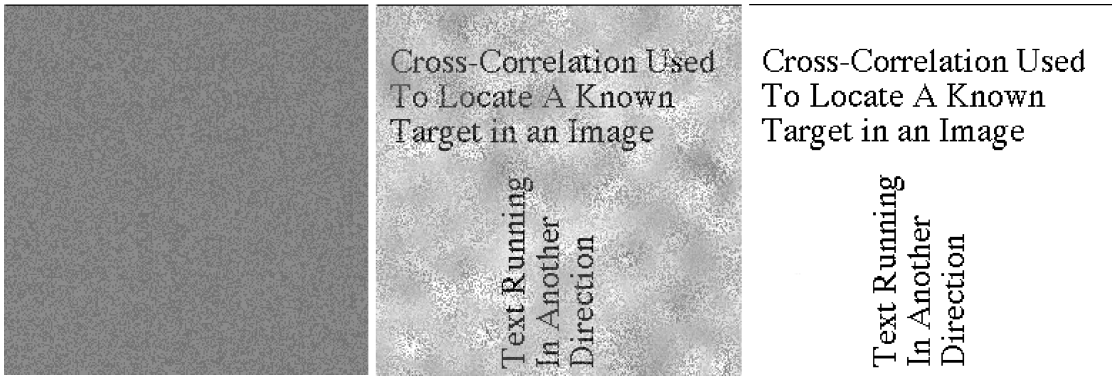

Fig. 28. The process of recall of one of three trained black-and-white images by an InBar-type CrossNet with 256×256 neural cells and connectivity parameter $M = 64$. The initial image (left panel) was obtained from the trained image (identical to the one shown in the right panel) by flipping 40% of randomly selected pixels. $\tau_0 = MR_LC_0 \lesssim RC_0$ is the effective time constant of intercell interaction. (C_0 is the dendrite wire capacitance per one synaptic plaquette.) [294].

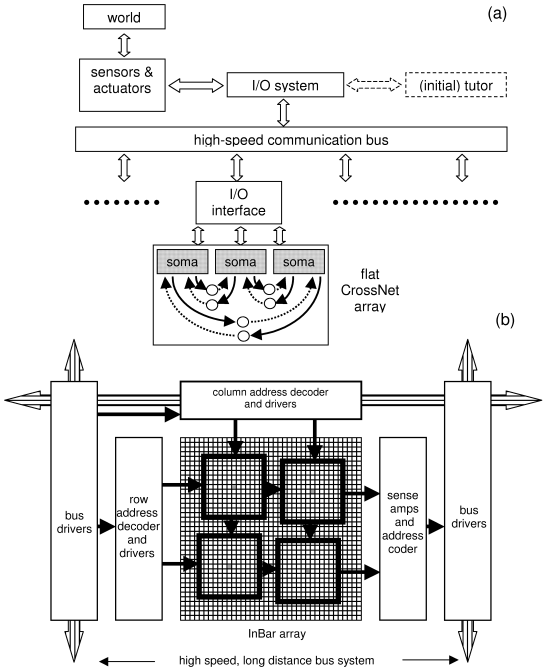

Fig. 29. (a) Possible hierarchical organization of a large-scale, CrossBar-based information system and (b) a simple scheme for incorporation of InBar matrices into the system fast (though relatively rare) communications of distant neural cells [294].

strate) [16] would be able, after a period of initial training by a dedicated external tutor (Fig. 29a), to learn directly from its interaction with environment. In this case one can speak of a "self-evolving" system.

If these expectations are confirmed, we may be able to revisit the initial dream of the neural network science of providing hardware means for reproducing the natural evolution of the neocortex on a much faster time scale. Such evolution may lead to self-development of such advanced features as system self-awareness (consciousness) and reasoning. If a substantial success along these lines materializes, it will have a strong impact not on the information technology, but also on the society as a whole.

5. Conclusions

I believe that the following main conclusions may be drawn from the materials presented in this chapter:

(i) Physics allows scaling of the silicon FETs, in their advanced (ultra-thin-channel, double-gate) form, to approximately 10 nm gate length, without an essential loss of performance. Further scaling, all way down to ∼ 5-nm-long channels, is also physically possible, but leads to an extremely high sensitivity of transistor characteristics (in particu-

[16] Such large-scale system would require a hierarchical organization (Fig. 29a) involving at least the means for fast signal transfer over long distances. Fortunately, for the InBar-type CrossNet with its regular location of somatic cell interfaces, such communication is easy to organize (Fig. 29b).

lar its gate threshold voltage V_t) to minute variations of geometric dimensions. This sensitivity will probably lead to unacceptable cost of fabrication facilities and, as a result, to the necessity of transfer to alternative electron devices in order to continue the Moore-Law-type exponential progress.

(ii) So far there is no single universal electronic device capable of replacing silicon FETs when their scaling down runs out of steam. The most natural candidates, single electron transistors, have the advantage of being remarkably material-insensitive, but suffer (as essentially all nanoscale devices) from low voltage gain and high sensitivity to single charge impurities in the dielectric environment of their islands. Nevertheless, there is a hope that advanced lithography may reach the level necessary for the fabrication of hybrid SET/FET memories. The major competition to this concept may come from NOVORAM memories based on crested tunnel barriers and 2T-cell memories, both using nanoscale FETs with ultra-thin channels.

(iii) For more advanced applications, the introduction of "CMOL" circuits, combining a CMOS stack, a few nanowiring layers, and a subsystem of molecular devices self-assembled on the wires seems unavoidable. If this concept is developed and implemented within the following 10 to 15 years, it will prevent the impending crisis of the exponential, Moore-Law-type progress of microelectronics, and extend this progress to virtually atomic dimensions.

(iv) The first application of the CMOL concept could be in the dedicated ultra-dense memory chips, as well as digital microprocessors with the molecular sub-system working as an embedded memory. However, CMOL circuits are the very natural hardware for the implementation of much more sophisticated architectures, in particular neuromorphic networks performing such advanced functions as pattern classification and feature detection, and possibly other human brain functions, at much higher speed. If such systems are eventually developed, the technological and social consequences of their practical introduction may be extremely significant.

Acknowledgments

The author is grateful to many colleagues for numerous discussions of the issues mentioned in this chapter. T. J. Walls and V. A. Sverdlov, have generously permitted the use of some results of our joint works [59, 61] prior to their publication, while A. Mayr, I. Muckra and O. Turel have granted the similar permission for work [294]. Yu. Pashkin has kindly provided original graphics for Fig. 10b. M. Macucci has brought to my attention some important recent publications and provided preprint [283]. Research at Stony Brook on certain topics discussed in this paper was supported in part by AFOSR, ARDA via ONR, DOE, NSF, and SRC. This paper is a substantially updated and extended version of the earlier review [99].

References

[1] International Technology Roadmap for Semiconductors, 2001 Edition, 2002 Update; available on the Web at `http://public.itrs.net/`.

[2] S. Sze, Physics of Semiconductor Devices, 2^{nd} ed. (Wiley, New York, 1981).

[3] Y. P. Tsividis, Operation and Modeling of the MOS Transistor (McGraw Hill, New York, 1987).

[4] R. F. Pierret, Semiconductor Device Fundamentals (Addison-Wesley, Reading, MA, 1996).

[5] S. Dmitrijev, Understanding Semiconductor Devices (Oxford Univ. Press, New York, 2000).

[6] D. J. Frank, R. H. Dennard, E. Nowak, P. M. Solomon, Y. Taur, and H.-S. P. Wong, Proc. IEEE 89 (2001) 259-288.

[7] H. Kawaura, T. Sakamoto, T. Baba, Y. Ochiai, J. Fujita, and J. Sone, IEEE Trans. on Electron Dev. 47 (2000) 856-859.

[8] H. Kawaura, T. Sakamoto, and T. Baba, Appl. Phys. Lett. 76 (2000) 3810-3812.

[9] R. Chau *et al.*, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2000) 45-48.

[10] H. Wakabayashi *et al.*, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2000) 49-52.

[11] M. Takahashi, T. Ohno, Y. Sakakibara, and K. Takayama, IEEE Trans. on Electron Dev. 48 (2001) 1380-1385.

[12] R. Chau, in: Proc. of Silicon Nanoelectronics Workshop (Honolulu, Hawaii, 2001) 57-60.

[13] B. Yu, H. Wang., A. Joshi, Q. Xiang, E. Ibok, and M.-R. Lin, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2001) 937-939.

[14] A. Hokazono *et al.*, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2002) 639-642.

[15] J. Appenzeller *et al.*, Appl. Phys. Lett. 77 (2000) 298-300.

[16] B. Doris *et al.*, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2002) 267-270.

[17] F. Balestra, S. Cristoloveanu, M. Benachir, and T. Elewa, IEEE Electron Dev. Lett. 8 (1987) 410-412.

[18] T. Tanaka, K. Suzuki, H. Horie, and T. Sugii, IEEE Electron Dev. Lett. 15 (1994) 386-388.

[19] I. Y. Yang, C. Vieri, A. Chandrakasan, and D. A. Antoniadis, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 1995) 877-880.

[20] H.-S. P. Wong, K. K. Chan, and Y. Taur, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 1997) 427-429.

[21] S. Tiwari, A. Kumar, and J. J. Welser, Int. J. High Speed Electron. Syst. 10 (2000) 231-245.

[22] V. W. C. Chan and P. C. H. Chan, IEEE Electron Dev. Lett. 22 (2001) 80-82.

[23] K. W. Guarini *et al.*, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2001) 425-428.

[24] M. Ieong *et al.*, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2001) 441-445.

[25] D. Hisamoto, T. Kaga, and E. Takeda, IEEE Trans. on Electron. Dev. 38 (1991)1419-1424.

[26] L. Guo, E. Leobandung, and S. Y. Chou, Science 275 (1997) 649-651.

[27] D. Hisamoto *et al.*, IEEE Trans. on Electron Dev. 47 (2000) 2320-2325.

[28] J. Kedzierski, P. Xuan, E. H. Anderson, J. Bokor, T.-J. King, and C. Hu, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2000) 57-60.

[29] Y.-K. Choi *et al.*, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2000) 421-424.

[30] X. Huang *et al.*, IEEE Trans. on Electron. Dev. 48 (2001) 880-884.

[31] T. Hiramoto, in: Proc. 2001 IEEE Int. SOI Conf. (IEEE Press, Piscataway, NJ, 2001) 8-10.

[32] J. Kedzierski *et al.,* in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2001) 437-440.

[33] B. Yu *et al.,* in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2002) 251-255.

[34] H. Takato *et al.*, IEEE Trans. on Electron Dev. 38 (1991) 573-577.

[35] M. Yang, C. L. Chang, M. Carroll and J. C. Sturm, IEEE Electron Dev. Lett. 20 (1999) 301-303.

[36] J. M. Hergenrother *et al.,* in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 1999) 75-78.

[37] J. M. Hergenrother *et al.*, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2001) 51-54.

[38] D. J. Frank, S. E. Laux, and M. V. Fischetti, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 1992) 553-556.

[39] K. Natori, J. Appl. Phys. 76 (1994) 4879-4890.

[40] M. S. Lundstrom, IEEE Electron Dev. Lett. 18 (1997) 361-363.

[41] F. G. Pikus and K. K. Likharev, Appl. Phys. Lett. 71 (1997) 3661-3663.

[42] S. Datta, F. Assad, and M. S. Lundstrom, Superlattices and Microstructures 23 (1998) 771-780.

[43] D. J. Frank, Y. Taur, and H.-S. P. Wong, IEEE Electron Dev. Lett. 19 (1998) 385-387.

[44] H.-S. P. Wong, D. J. Frank, and P. M. Solomon, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 1998) 407-410.

[45] Y. Naveh and K. K. Likharev, IEEE Electron Dev. Lett. 21 (2000) 242-244.

[46] S. Cristoloveanu, T. Ernst, D. Munteanu, and T. Ouisse, Int. J. of High Speed Electron. and Syst. 10 (2000) 217-230.

[47] F. Assad, Z. Ren, D. Vasileska, S. Datta, and M. Lundstrom, IEEE Trans. on Electron Dev. 47 (2000) 232-240.

[48] J. G. Fossum, Z. Ren, K. Kim, and M. Lundstrom, Superlattices and Microstructures 28, (2000) 525-530.

[49] M. S. Lundstrom, IEEE Electron Dev. Lett. 22 (2001) 293-295.

[50] V. Sverdlov, Y. Naveh, and K. Likharev, in: Proc. 2001 IEEE Int. SOI Conf. (IEEE Press, Piscataway, NJ, 2001) 151-152.

[51] P. M. Solomon, and S. E. Laux, in: Tech. Dig. of IEDM (IEEE Press, Piscataway, NJ, 2001) 95-98.

[52] Z. Ren, R. Venugopal, S. Datta, and M. Lundstrom, in: Tech. Dig. of IEDM (IEEE Press, Piscataway, NJ, 2001) 107-110.

[53] A. Svizhenko, M. P. Anantram, T. R. Govindan, and R. Venugopal, J. Appl. Phys. 91 (2002) 2343-2354.

[54] M. Lundstrom, and Z. B. Ren, IEEE Trans. on Electron Dev. (2002) 133-141 .

[55] R. Venugopal, Z. Ren, S. Datta, and M. S. Lundstrom, J. Appl. Phys. 92 (2002) 3730-3739.

[56] J. H. Rhew, Z. B. Ren, and M. S. Lundstrom, Solid-State Electronics 46 (2002) 1899-1906.

[57] J. Wang and M. Lundstrom, in: Tech. Dig. of IEDM (IEEE Press, Piscataway, NJ, 2002) 707-710.

[58] S. E. Laux, A. Kumar, and M. V. Fischetti, in: Tech. Dig. of IEDM (IEEE Press, Piscataway, NJ, 2002) 715-718.

[59] V. A. Sverdlov, T. J. Walls, and K. K. Likharev, Nanoscale silicon MOSFETs: A theoretical study, accepted for publication in IEEE Trans. on Electron Dev. (2003).

[60] A. Svizhenko and M. P. Anantram, Scattering in nanotransistors: An analytical study, preprint (2002).

[61] T. J. Walls, V. A. Sverdlov, and K. K. Likharev, MOSFETs below 10 nm: Quantum Theory, Report at NanoMES'03 (Tempe, AZ, Feb. 2003), to be published in Physica E (2003).

[62] D. Esseni *et al.*, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2001) 445-449.

[63] S. Takagi, A. Torumi, M. Iwase, and H. Tango, IEEE Trans. on Electron Dev. 41 (1994) 2357-2362.

[64] M. Fischetti, S. E. Laux, and E. Crabbe, J. Appl. Phys. 78 (1995) 1058-1087.

[65] T. Sakamoto, H. Kawaura, T. Baba, and T. Iizuka, Appl. Phys. Lett. 76 (2000) 2618-2620.

[66] P. M. Solomon (solomon@watson.ibm.com), On contacts to small semiconductor devices, unpublished (1998).

[67] V. Sverdlov, X. Oriols, and K. Likharev, IEEE Trans. on Nanotechnology 2 (2002) 59-63.

[68] K. K. Young,. IEEE Trans. on Electron Dev. 36 (1989) 504-506.

[69] K. Suzuki, T. Tanaka, Y. Tosaka, H. Hoire, and Y. Arimoto, IEEE Trans. on Electron Dev. 40 (1993) 2326-2329.

[70] S. P. Tang, R. M. Wallace, A. Seabaugh, D. King-Smith, Appl. Surf. Sci. 135 (1998) 137-142.

[71] D. A. Muller, T. Sorsch, S. Moccio, F. H. Baumann, K. Evans-Lutterodt, and G. Timp, Nature, 399 (1999) 758-761.

[72] M. S. Krishnan, L. Chang, T.-J. King, J. Bokor, and C. Hu, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 1999) 241-244.

[73] S. Song *et al.*, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2001) 55-58.

[74] J. P. Colinge, Silicon-On-Insulator Technology (Kluwer, Dordrecht, 1991).

[75] S. Cristoloveanu, Int. J. of High Speed Electron. Syst. 10 (2000) 129-130.

[76] F. Wittman, H. Gossner, and I. Eisele, J. of Mat. Sci.: Materials in Electronics 6 (1995) 336-340.

[77] L. Wong, H. Qu, J.-M. Kuo, and K. K. Chin, IEEE Trans. on Electron Dev. 46 (1999) 865-870.

[78] K. Nakazato, K. Itoh, H. Mizuta, and H. Ahmed, Electron Lett. 35 (1999) 848-850.

[79] S. Xu, K. P. Gan, P. D. Foo, and Y. Liu, IEEE Electron Dev. Lett. 21 (2000) 176-178.

[80] F. Allibert, A. Zaslavsky, and S. Cristoloveanu, in: Proc. of 2001 IEEE Int. SOI Conf. (IEEE Press, Piscataway, NJ, 2001) 149-150.

[81] Q. Lu *et al.*, IEEE Electron Dev. Lett. 22 (2001) 324-326.

[82] L. Kang *et al.*, IEEE Electron Dev. Lett. 21 (2000) 181-183.

[83] W. J. Qi *et al.*, Appl. Phys. Lett. 77 (2000) 1704-1706.

[84] E. P. Gusev *et al.*, in: Tech. Dig. IEDM (IEEE Press, Piscataway, NJ, 2001) 451-454.

[85] L. D. Landau and E. M. Lifshitz, Quantum Mechanics (Non-Relativistic Theory), 3^{rd} ed. (Pergamon Press, Oxford, 1977) 184-185.

[86] J. A. López-Villanueva, P. Cartujo-Castinello, F. Gámiz, J. Banqueri, and A. J. Palma, IEEE Trans. on Electron Dev. 47 (2000) 141-146.

[87] R. Sasajima, K. Fujimaru, and H. Matsumura, Appl. Phys. Lett. 74 (1999) 3215-3217.

[88] C.-K. Huang, W. E. Zhang, and C. H. Yang, IEEE Trans. on Electron Dev. 45 (1998) 842-848.

[89] C. Wang, J. P. Snyder, and J. R. Tucker, Appl. Phys. Lett. 65 (1999) 618-620.

[90] P. P. Gelsinger, in: ISSCC Tech. Dig. (IEEE Press, Piscataway, NJ, 2001) 22-25.

[91] D. J. Frank, IBM J. of Res. Devel. 46 (2002) 235-244.

[92] S. J. Tans, A. R. M. Verschueren, and C. Dekker, Nature 393 (1998) 49-52.

[93] C. Zhou, J. Kong, and H. Dai, Appl. Phys. Lett. 76 (2000) 1597-1599.

[94] R. Martel, H.-S. P. Wong, K. Chan, and P. Avouris, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2001) 159-161.

[95] S. Rosenblatt, Y. Yaish, J. Park, J. Gore, V. Sazonova, and P. L. McEuen, Nano Letters 2 (2002) 869-872.

[96] R. Martel, T. Schmidt, H. R. Shea, T. Hertel, and P. Avouris, Appl. Phys. Lett. 73 (1998) 2447-2449.

[97] W. Roush, IEEE Spectrum 38 (2001) 25-26.

[98] R. F. Service, Science 293 (2001) 785-786.

[99] K. K. Likharev, in: H. Morkoç, ed., Advanced Semiconductor and Organic Nano-Techniques, Part I (Elsevier, San Diego, 2003) 239-302.

[100] B. Altshuler, P. Lee, and R. Webb, eds., Mesoscopic Phenomena in Solids (Elsevier, Amsterdam, 1991).

[101] Y. Imry, Introduction to Mesoscopic Physics (Oxford Univ. Press, Oxford, 1997).

[102] S. Datta, Electronic Transport in Mesoscopic Systems (Cambridge Univ. Press. Cambridge, MA, 1997).

[103] S. Datta, Superlattices and Microstructures 6 (1989) 83-93.

[104] F. Capasso, ed., Physics of Quantum Electron Devices (Springer, Berlin, 1990).

[105] A. J. Holden, in: K. E. Singer, ed., Gallium Arsenide and Related Compounds (IOP Publishing, Bristol, UK, 1990) 1-78.

[106] T. C. L. G. Sollner, E. R. Brown, W. D. Goodhue, and H. Q. Lee, in: F. Capasso, ed., Physics of Quantum Electron Devices (Springer, Berlin, 1990) 147-158.

[107] L. L. Chang, E. E. Mendez, and C. Tejedor, eds., Resonant Tunneling in Semiconductors (Plenum, New York, 1992).

[108] P. E. Thompson *et al.*, Thin Solid Films 380 (2000) 145-150.

[109] M. Tinkham, Introduction to Superconductivity, 2^{nd} ed. (McGraw-Hill, New York, 1996).

[110] K. Likharev, Dynamics of Josephson Junctions and Circuits (Gordon and Breach, New York, 1986).

[111] D. K. Brock, E. K. Track, and J. M. Rowell, IEEE Spectrum 37 (2000) 40-46.

[112] P. Bunyk, K. Likharev, and D. Zinoviev, Int. J. on High Speed Electronics and Systems 11 (2001) 257-305.

[113] M. A. Nielsen and I. L. Chuang, Quantum Computation and Quantum Information (Cambridge Univ. Press, Cambridge, UK, 2000).

[114] K. K. Likharev, IBM J. Res. Devel. 32 (1988) 144-158.

[115] D. V. Averin and K. K. Likharev, in: B. Altshuler, P. Lee, and R. Webb, eds., Mesoscopic Phenomena in Solids (Elsevier, Amsterdam, 1991) 173-271.

[116] M. H. Devoret and H. Grabert, eds., Single Charge Tunneling (Plenum, New York, 1992).

[117] K. K. Likharev, Proc. IEEE 87 (1999) 606-632.

[118] J. Lambe and R. C. Jaklevic, Phys. Rev. Lett. 22 (1969) 1371-1375.

[119] I. O. Kulik, and R. I. Shekhter, Sov. Phys. - JETP 41 (1975) 308-316.

[120] P. Lafarge *et al.*, Z. Phys. B 85 (1991) 327-332.

[121] Sh. Farhangfar *et al.* J. Low Temp. Phys. 108 (1997) 191-215.

[122] J. P. Pekola *et al.* J. Low Temp. Phys. 128 (2002) 263-269.

[123] D. V. Averin and K. K. Likharev, J. Low Temp. Phys. 62 (1986) 345-372.

[124] K. K. Likharev, IEEE Trans. on Magn. 23 (1987) 1142-1145.

[125] T. A. Fulton, and G. D. Dolan, Phys. Rev. Lett. 59 (1987) 109-112.

[126] Yu. A. Pashkin, Y. Nakamura, and J. S. Tsai, Appl. Phys. Lett. 76 (2000) 2256-2258.

[127] G. Zimmerli, R. L. Kautz, and J. M. Martinis, Appl. Phys. Lett. 61 (1992) 2616-2618.

[128] R. A. Smith, and H. Ahmed, Appl. Phys. Lett. 71 (1997) 3838-3840.

[129] Y. Ono, Y. Takahashi, K. Yamazaki, M. Nagase, H. Namatsu, K. Kurihara, and K. Murase, Appl. Phys. Lett. 76 (2000) 3121-3123.

[130] T. A. Fulton, P. L. Gammel, and L. N. Dunkleberger, Phys. Rev. Lett. 67 (1991) 3148-3151.

[131] D. V. Averin and K. K. Likharev, in: M. H. Devoret and H. Grabert, eds., Single Charge Tunneling (Plenum, New York, 1992) 311-322.

[132] L. J. Geerligs *et al.* Phys. Rev. Lett. 64 (1990) 2691-2694.

[133] K. K. Likharev, N. S. Bakhvalov, G. S. Kazacha, and S. I. Serdyukova, IEEE Trans. on Magn. 25 (1989) 1436-1439.

[134] K. K. Likharev and K. A. Matsuoka, Appl. Phys. Lett. 67 (1995) 3037-3039.

[135] D. V. Averin and A. A. Odintsov, Phys. Lett. A 140 (1989) 251-255.

[136] P. D. Dresselhaus, L. Ji, S. Y. Han, J. E. Lukens, and K. K. Likharev, Phys. Rev. Lett. 72 (1994) 3226-3229.

[137] L. Ji, P. D. Dresselhaus, S. Y. Han, K. Lin, W. Zheng, and J. E. Lukens, J. Vac. Sci. Technol. B 12 (1994) 3619-3622.

[138] K. A. Matsuoka, K. K. Likharev, P. D., Dresselhaus, L. Ji, S. Y. Han and J. E. Lukens, J. Appl. Phys. 81 (1997) 2269-2281.

[139] K. K. Likharev and A. N. Korotkov, Science 273 (1996) 763-765.

[140] A. N. Korotkov and K. K. Likharev, J. Appl. Phys. 84 (1998) 6114-6126.

[141] A. O. Orlov, R. K. Kummamuru, R. Ramasubramaniam, T. Toth, C. S. Lent, G. H. Bernstein, and G. L. Snider, Appl. Phys. Lett. 78 (2001) 1625-1627.

[142] D. V. Averin and A. N. Korotkov, J. Low Temp. Phys. 80 (1991) 173-185.

[143] D. V. Averin, A. N. Korotkov and K.K. Likharev, Phys. Rev. B 44 (1991) 6199-6211.

[144] L. P. Kouwenhoven *et al.*, in: L. Sohn *et al.*, eds., Mesoscopic Electron Transfer (Kluwer, Dordrecht, 1997) 105-215.

[145] K. Matsumoto, M. Ishii, K. Segawa, Y. Oka, B. J. Vartanian, and J. S. Harris, Appl. Phys. Lett. 68 (1996) 34-36.

[146] K. Matsumoto, Y. Gotoh, T. Maeda, J. A. Dagata, and J. S. Harris, Appl. Phys. Lett. 76 (2000) 239-241.

[147] H. W. Ch. Postma, T. Teepen, Z. Yao, M. Grifoni, and C. Dekker, Science 293 (2001) 76-79.

[148] C. Thelander *et al.*, Appl. Phys. Lett. 79 (2001) 2106-2108.

[149] J. B. Cui, M. Burghard, and M. Kern, Nano Letters 2 (2002) 117-120.

[150] Y. Takahashi *et al.*, Electron. Lett. 31 (1995) 136-138.

[151] Y. Ono, Y. Takahashi, K. Yamazaki, M. Nagase, H. Namatsu, K. Kurihara, and K. Murase, IEEE Trans on Electron. Dev. 47 (2000) 147-153.

[152] D. H. Kim, J. D. Lee, and B.-G. Park, Jpn. J. Appl. Phys. (pt. 1) 39 (2000) 2329-2333.

[153] K. Uchida, J. Koga, R. Ohba, and A. Torumi, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2000) 863-865.

[154] M. Saitoh, N. Takahashi, H. Ishikuro, and T. Hiramoto, Jpn. J. Appl. Phys. (pt. 1) 40 (2001) 2010-2012.

[155] N. J. Tao, Phys. Rev. Lett. 76 (1996) 4066-4069.

[156] E. S. Soldatov *et al.*, JETP Lett. 64 (1996) 556-558.

[157] M. A. Reed, C. Zhou, C. J. Miller, T. P. Burgin, and J. M. Tour, Science 278 (1997) 252-254.

[158] W. Han *et al.*, J. Phys. Chem. B 101 (1997) 10719-10725.

[159] C. P. Collier *et al.*, Science 285 (1999) 391-394.

[160] C. Kergueris *et al.*, Phys. Rev. B 59 (1999) 12505-12513.

[161] J. Chen, M. A. Reed, A. M. Rawlett, and J. M. Tour, Science 286 (1999) 1550-1552.

[162] D. Porath, A. Bezryadin, S. de Vries, and C. Dekker, Nature 403 (2000) 635-638.

[163] C. P. Collier *et al.*, Science 289 (2000) 1172-1175.

[164] D. I. Gittins, D. Bethell, D. J. Schiffrin, and R. J. Nichols, Nature 408 (2000) 67-69.

[165] H. Watanabe, C. Manabe, T. Shigematsu, and M. Shimuzu, Appl. Phys. Lett. 78 (2001) 2928-2930.

[166] Z. J. Donhauser *et al.*, Science 292 (2001) 2303-2306.

[167] H. Xe *et al.*, J. Am. Chem. Soc. 123 (2001) 7730-7731.

[168] X. D. Cui *et al.*, Science 294 (2002) 571-574.

[169] S. P. Gubin *et al.*, Nanotechnology 13 (2002) 185-194.

[170] J. Reichert, R. Ochs, D. Beckman, H. B. Weber, M. Mayor, and H. v. Löhneysen, Phys. Rev. Lett. 88 (2002) 176804 1-4.

[171] N. B. Zhitenev, H. Meng, and Z. Bao, Phys. Rev. Lett. 88 (2002) 226801 1-4.

[172] J. Park *et al.*, Nature 417 (2002) 722-725.

[173] W. Liang, M. P. Shores, M. Bockrath, J. R. Long, and H. Park, Nature 417 (2002) 725-729.

[174] H. B. Weber *et al.*, Chem. Physics 281 (2002) 113-125.

[175] J. Chen and M. Reed, Chem. Physics 281 (2002) 127-145.

[176] R. H. M. Smit *et al.*, Nature 419 (2002) 906-909.

[177] D. B. Janes, H. Halimun, J. Choi, and S. Burns, Metal-molecule-metal; structures with pre-fabricated contacts, Report at the 6^{th} Conf. on Molecular-scale Electronics (Key West, FL, Dec. 2002), to be published by New York Acad. Sci. (2003).

[178] M. Reed, Proc. IEEE 87 (1999) 652-705.

[179] M. Di Ventra, S. T. Pantelides, and N. D. Lang, Phys. Rev. Lett. 84 (2000) 979-982.

[180] V. V. Shorokhov, P. Johanson, and E. S. Soldatov, J. Appl. Phys. 91 (2002) 3049-3053.

[181] G. Cuniberti, L. Craco, D. Porath, and C. Dekker, Phys. Rev. B 65 (2002) 241314 (R) 1-4.

[182] G. Cuniberti, J. Yi, and M. Porto, Appl. Phys. Lett. 81 (2002) 850-852.

[183] N. Zimbovskaya and G. Gumbs, Appl. Phys. Lett. 81 (2002) 1518-1520.

[184] P. E. Kornilovitch, A. M. Bratkovsky, and R. S. Williams, Phys. Rev. B 66 (2002) 165436 1-12.

[185] D. L. Allara *et al.*, Ann. of New York Acad. Sci. 852 (1998) 349-370.

[186] E. H. Nicollian and J. R. Brews, MOS physics and technology (Wiley, New York, 1982).

[187] A. N. Korotkov, Appl. Phys. Lett. 72 (1998) 3226-3228.

[188] Yu. A. Pashkin, Y. Nakamura, and J. S. Tsai, Appl. Phys. Lett. 74 (1999) 132-134.

[189] A. M. Ionescu, M. J. Declercq, K. Banerjee, and J. Gautier, in: Proc. of ACM IEEE Design Automation Conference (ACM, New York, 2002) 88-93.

[190] W. Zheng, J. R. Friedman, D. V. Averin, S. Han, and J. E. Lukens, Solid State Commun. 108 (1998) 839-843.

[191] K. E. Nagaev, Phys. Lett. A 169 (1992) 103-107.

[192] K. E. Nagaev, Phys. Rev. B 52 (1995) 4740-4743.

[193] Y. Naveh, D. V. Averin, and K. K. Likharev, Phys. Rev. B 58 (1998) 15371-15374.

[194] V. K. Sverdlov, A. N. Korotkov, and K. K. Likharev, Phys. Rev. B 63 (2000) R081302 1-4.

[195] Y. A. Kinkhabwala, V. A. Sverdlov, A. N. Korotkov, and K. Likharev, Shot noise at hopping: A numerical study, ArXiv preprint cond-mat/0302445, submitted to Phys. Rev. B (2003).

[196] D. M. Kaplan, V. A. Sverdlov, and K. K. Likharev, Coulomb gap, Coulomb blockade, and dynamic activation energy in frustrated single-electron arrays, ArXiv preprint cond-mat/0303439, submitted to Phys. Rev. B (2003).

[197] Y. Pekola, in: H. Morkoç, ed., Advanced Semiconductor and Organic Nano-Techniques, (Academic Press, San Diego, 2003), pt. 1.

[198] K. Likharev, Single-electron devices, to be published in: Encyclopedia of Nanoscience and Nanotechnology (American Scientific Publishers, Stevenson Ranch, CA, 2003).

[199] A. N. Korotkov, D. V. Averin, K. K. Likharev, and S. A. Vasenko, in: H. Koch and H. Lübbig, eds., Single Electron Tunneling and Mesoscopic Devices (Springer, Berlin, 1992) 45-60.

[200] A. N. Korotkov, Phys. Rev. B 49 (1994) 10381-10392.

[201] V. A. Krupenin, D. E. Presnov, A. B. Zorin, and J. Niemeyer, Physica B 284 (2000) 1800-1801.

[202] R. J. Schoelkopf, P. Wahlgren, A. A. Kozhevnikov, P. Delsing, and D. E. Prober, Science 280 (1998) 1238-1242.

[203] A. Aassime, G. Johansson, G. Wendin, R. J. Schoelkopf, and P. Delsing, Phys. Rev. Lett. 86 (2001) 3376-3379.

[204] A. N. Korotkov and M. A. Paalanen, Appl. Phys. Lett. 74 (1999) 4052-4054.

[205] B. Prince, Semiconductor Memories, 2^{nd} ed. (Wiley, Chichester, UK, 1991).

[206] W. D. Brown and J. E. Brewer, Nonvolatile Semiconductor Memory Technology (IEEE Press, Piscataway, NJ, 1998).

[207] S. Tiwari, F. Rana, H. Hanafi, A. Hartstein, E. F. Crabbé, and K. Chan, Appl. Phys. Lett. 68 (1996) 1377-1379.

[208] H. I. Hanafi, S. Tiwari, and I. Khan, IEEE Trans. on Electron. Dev. 43 (1996) 1553-1558.

[209] Y. Shi, K. Saito, H. Ishikuro, and T. Hiramoto, J. Appl. Phys. 84 (1998) 2358-2360.

[210] Y. C. King, T. J. King, and C. M. Hu, IEEE Electron. Device Lett. 20 (1999) 409-411.

[211] I. Kim, S. Han, K. Han, J. Lee, and H. Shin, IEEE Electron. Device Lett. 20 (1999) 630-631.

[212] K. Han, I. Kim, and H. Shin, IEEE Trans. on Electron Devices 48 (2001) 874-879.

[213] R. Ohba, N. Sugiyama, J. Koga, K. Uchida, and A. Torumi, Jpn. J. Appl. Phys. (pt. 1) 39 (2000) 989-993.

[214] A. Fernandes *et al.*, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2001) 155-158.

[215] T. Ishii, T. Osabe, T. Mine, F. Murai, and K. Yano, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2001) 305-308.

[216] A. Nakajima, T. Futatsugi, K. Kosemura, T. Fukano, and N. Yokoyama, Appl. Phys. Lett. 70 (1997) 1742-1744.

[217] J. J. Welser, S. Tiwari, S. Rishton, K. Y. Lee, and Y. Lee, IEEE Electron Device Lett. 18 (1997) 278-280.

[218] T. Tsutsumi, K. Ishii, E. Suzuki, H. Hiroshima, M. Yamanaka, I. Sakata, S. Kanemaru, S. Hazra, T. Maeda, and K. Tomizawa, Electron. Lett. 36 (2000) 1322-1323.

[219] K. Yano, T. Ishii, T. Sano, T. Mine, F. Murai, T. Hashimoto, T. Kobayashi, T. Kure, and K. Seki, Proc. IEEE 87 (1999) 633-651.

[220] K. Yano, T. Ishii, T. Sano, T. Mine, F. Murai, T. Kure, and K. Seki, in:ISSCC'98 Digest of Technical Papers (IEEE Press, New York, 1998) 344-345.

[221] K. K. Likharev, and A. N. Korotkov, in: Proc. of the 1995 Int. Semicond. Device Res. Symp. (Univ. of Virginia, Charlottesville, VA, 1995) 355-358.

[222] C. D. Chen, Y. Nakamura, and J. S. Tsai, Appl. Phys. Lett. 71 (1997) 2038-2040.

[223] A. N. Korotkov, J. Appl. Phys. 92 (2002) 7291-7295.

[224] K. K. Likharev, Appl. Phys. Lett. 73 (1998) 2137-2139.

[225] A. Korotkov and K. Likharev, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 1999) 223-226.

[226] K. K. Likharev, IEEE Circuits and Devices Mag. 16 (2000) 16-21.

[227] H. Shichijyo *et al.*, in: Ext. Abstr. of the 16^{th} Int. Conf. on Solid State Devices and Materials (Business Center for Academic Societies, Tokyo, 1984) 265-267.

[228] N. J. Stone and H. Ahmed, Appl. Phys. Lett. 73 (1998) 2134-2136.

[229] K. Nakazato, and H. Ahmed, Appl. Phys. Lett. 66 (1995) 3170-3172.

[230] Z. A. K. Durrani, A. C. Irvine, and H. Ahmed, Appl. Phys. Lett. 74 (1999) 1293-1295.

[231] Z. A. K. Durrani, A. C. Irvine, and H. Ahmed, IEEE Trans. on Electron. Devices 47 (2000) 2334-2339.

[232] A. C. Irvine, Z. A. K. Durrani, and H. Ahmed, J. Appl. Phys. 87, 8594-8603 (2000).

[233] H.–O. Müller, K. Katayama, and H. Mizuta, J. Appl. Phys. 84 (1998) 5603-5609.

[234] T. Osabe, T. Ishii, T. Mine, F. Murai, and K. Yano, in: Techn. Digest IEDM (IEEE Press, Piscataway, NJ, 2000) 301-304.

[235] K. Nakazato, K. Itoh, H. Mizuta, and H. Ahmed, Electron Lett. 35 (1999) 848-850.

[236] K. Nakazato, K. Itoh, H. Ahmed, H. Mizuta, T. Kisu, M. Kato, and T. Sakata, in: Dig. of Techn. Papers of ISSCC (IEEE Press, Piscataway, NJ, 2000) 132-133.

[237] J. H. Yi, W. S. Kim, S. Song, Y. Khang, H.-I. Kim, J. H. Choi, H. H. Lim, N. I. Lee, K. Fujihara, H.-K. Kang, J. T. Moon, and M. Y. Lee, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2001) 787-790.

[238] J. F. Scott, Ferroelectric Memories (Springer, New York, 2000).

[239] S. Tehrani, B. Engel, J. M. Slaughter, E. Chen, M. DeHerrera, M. Durlam, P. Naji, P. Whig, J. Janesky, and J. Calder, IEEE Trans. on Magn. 36 (2000) 2752-2757.

[240] K. Inomata, IEICE Trans. on Electronics E84C (2001) 740-746.

[241] S. Lai and T. Lowrey, in: Techn. Dig. IEDM (IEEE Press, Piscataway, NJ, 2001), 803-806.

[242] Ovonic unified memory, available on the Web at `http://www.ovonyx.com/ovonyxtech.html`.

[243] R. K. Kummamuru *et al.*, Appl. Phys. Lett. 81 (2002) 1332-1334.

[244] C. Kothandaramann, S. K. Iyer, and S. S. Iyer, IEEE Electron Dev. Lett. 23 (2002) 523-525.

[245] HP announces breakthroughs in molecular electronics, press release, available on the Web at `http://www.hp.com/hpinfo/newsroom/press/09sep02a.htm` (2002).

[246] M. J. Yoo *et al.*, Science 276 (1997) 579-582.

[247] J. R. Tucker, J. Appl. Phys. 72 (1992) 4399-4413.

[248] A. N. Korotkov, R. H. Chen, and K. K. Likharev, J. Appl. Phys. 78 (1995) 2520-2530.

[249] R. H. Chen, A. N. Korotkov, and K. K. Likharev, Appl. Phys. Lett. 68 (1996) 1954-1956.

[250] H. Iwamura, M. Akazawa, and Y. Amemiya, IEICE Trans. on Electron. E81-C (1998) 42-48.

[251] R. H. Chen and K. K. Likharev, Appl. Phys. Lett. 72 (1998) 61-63.

[252] K. K. Likharev, and V. K. Semenov, in: Ext. Abstr. of Int. Supercond. Electronics Conf., (Tokyo, 1987) 128-131.

[253] S. V. Vyshenskii, S. P. Polonsky, and K. K. Likharev, Single electron logic circuits, unpublished (1990), described in detail in Ref. 131.

[254] P. Bakshi, D. A. Broido, and K. Kempa, J. Appl. Phys. 70 (1991) 5150-5152.

[255] C. S. Lent, P. D. Tougaw, W. Porod, and G. Bernstein, Nanotechnology 4 (1993) 49-57.

[256] K. Nomoto, R. Ugajin, T. Suzuki and I. Hase, Electron. Lett. 29, (1993) 1380-1381.

[257] D. V. Averin, L. F. Register, K. K. Likharev, and K. Hess, Appl. Phys. Lett. 64 (1994) 126-128.

[258] P. D. Tougaw and C. S. Lent, J. Appl. Phys. 75 (1994) 1818-1825.

[259] S. Bandyopadhyay, B. Das, and A. E. Miller, Nanotechnology 4 (1994) 113-133.

[260] A. N. Korotkov, Appl. Phys. Lett. 67 (1995) 2412-2414.

[261] M. G. Ancona, J. Appl. Phys. 78 (1995) 3311-3314.

[262] K. Nomoto, R. Ugajin, T. Suzuki and I. Hase, J. Appl. Phys. 79 (1996) 291-300.

[263] C. S. Lent and P. D. Tougaw, J. Appl. Phys. 80 (1996) 4722-4736.

[264] S. Bandyopadhyay and V. Roychowdhury, Jpn. J. Appl. Phys. (pt. 1) 35 (1996) 3350-3362.

[265] A. V. Krasheninnikov and L. A. Openov, JETP Lett. 64 (1996) 231-218.

[266] S. N. Molotkov and S. S. Nazin, JETP 83 (1996) 794-802.

[267] C. S. Lent and P. D. Tougaw, Proc. IEEE 85 (1997) 541-557.

[268] N.-J. Wu, N. Asahi, and Y. Amemiya, Jpn. J. Appl. Phys. 36 (1997) 2621-2627.

[269] N. Asahi, M. Akazawa, and Y. Amemiya, IEEE Trans. on Electron. Dev. 44 (1997) 1109-1116.

[270] A. M. Bychkov, L. A. Openov, and I. A. Semenihin, JETP Lett. 66 (1997) 298-303.

[271] Y. Fu and M. Willander, J. Appl. Phys. 83 (1998) 3186-3191.

[272] M. P. Anantram and V. P. Roychowdhury, J. Appl. Phys. 85 (1999) 1622-1625.

[273] M. Governale, M. Macucci, G. Iannaccone, and C. Ungarelli, J. Appl. Phys. 85 (1999) 2962-2971.

[274] A. Gin, S. Williams, H. Y. Meng, and P. D. Tougaw, J. Appl. Phys. 85 (1999) 3713-3720.

[275] A. Gin, P. D. Tougaw, S. Willams, J. Appl. Phys. 85 (1999) 8281-8286.

[276] M. Girlanda, M. Governale, M. Macucci, and G. Iannaccone, Appl. Phys. Lett. 75 (1999) 3198-3200.

[277] C. Ungarelli, S. Francaviglia, and M. Macucci, and G. Iannaccone, J. Appl. Phys. 87 (2000) 7320-7325.

[278] J. R. Pasky, L. Henry, and P. D. Tougaw, J. Appl. Phys. 87 (2000) 8604-8609.

[279] M. T. Niemier and P. M. Kogge, Int. J. of Circuit Theory and Appl. 29 (2001) 49-62.

[280] L. Bonci, G. Iannaccone, and M. Macucci, J. Appl. Phys. 89 (2001) 6435-6443.

[281] J. Timler and C. S. Lent, J. Appl. Phys. 91 (2002) 823-831.

[282] L. Bonci, M. Gattobigio, G. Iannaccone, and M. Macucci, J. Appl. Phys. 92 (3169) 3169-3178.

[283] M. Macucci *et al.*, Critical assessment of the QCA architecture as a viable alternative to large scale integration, preprint (2003).

[284] R. A. Kiehl and T. Ohshima, Appl. Phys. Lett. 67 (1995) 2494-2496.

[285] T. Ohshima, and R. A. Kiehl, J. Appl. Phys. 80 (1996) 912-923.

[286] T. Ohshima, Appl. Phys. Lett. 69 (1996) 4059-4061

[287] F. Y. Liu, F. T. An, and R. A. Kiehl, Appl. Phys. Lett. 74 (1999) 4040-4042.

[288] P. Delsing, L. S. Kuzmin, K. K. Likharev, and T. Claeson, Phys. Rev. Lett. 63 (1989) 1861-1865.

[289] Y. Wada, Ann. New York Acad. Sci. 960 (2002) 39-61.

[290] S. Zankovych, T. Hoffmann, J. Seekamp, J. U. Bruch, and C. M. S. Torres, Nanotechnology 12 (2001) 91-95.

[291] J. M. Tour, M. Kozaki, and J. M. Seminario, J. Am. Phys. Soc. 120 (1998) 8486-8493.

[292] S. Fölling, Ö. Türel, and K. K. Likharev, in: Proc. of the 2001 Int. Joint Conf. on Neural Networks (Int. Neural Network Society, Mount Royal, NJ, 2001) 216-221.

[293] Ö. Türel, and K. K. Likharev, Int. J. of Circuit Theory and Appl. 31 (2003) 37-54.

[294] K. Likharev, A. Mayr, I. Muckra, Ö. Türel, CrossNets: Fault-tolerant architectures for molecular electronic circuits, report at the 6^{th} Conf. On Molecular-Scale Electronics (Key West, Dec. 2002), to be published by the New York Acad. Sci. (2003).

[295] D. J. Amit, Modeling Brain Function (Cambridge U. Press, Cambridge, UK, 1989).

[296] J. Hertz, A. Krogh, R. G. Palmer, Introduction to the Theory of Neural Computation. (Perseus, Cambridge, MA, 1991).

[297] P. S. Churchland and T. J. Sejnowski, The Computational Brain (MIT Press, Cambridge, MA, 1992).

[298] L. Fausett, Fundamentals of Neural Networks (Prentice Hall, Upper Saddle River, NJ, 1994).

[299] B. Müller, J. Reinhard, M. T. Strickland, Neural Networks, 2^{nd} ed. (Springer, Berlin, 1995).

[300] M. H. Hassoun, Fundamentals of Artificial Neural Networks (MIT Press, Cambridge, MA, 1995).

[301] S. Haykin, Neural networks (Prentice Hall, Upper Saddle River, NJ, 1999).

[302] W. S. Sarle, Neural networks: FAQ, papers, and various other things, available for ftp download from ftp.sas.com/pub/neural/FAQ.html.zip, 2002.

[303] V. B. Mountcastle, Perceptual Neuroscience. The cerebral cortex (Harvard U. Press, Cambridge, MA, 1998).

[304] V. Braitenberg and A. Schüz, Cortex: Statistics and Geometry of Neuronal Connectivity, 2^{nd} ed. (Springer, Berlin, 1998).

[305] M. Akazawa and Y. Ameniya, Appl. Phys. Lett. 70 (1997) 670-673.

[306] D. Gosztola, M. P. Niemczyk, W. Svec, A. S. Lukas, and M. R. Wasielewski, J. Phys. Chem. A 104 (2000) 6545-6551.

[307] H. D. Sikes *et al.*, Science 291 (2001) 1519-1523.

[308] H. McNally, D. B. Jones, B. Kasibhatla, C. P. Kubiak, Superlattices and Microstructures 31 (2002) 239-245.

[309] E. G. Emiroglu, Z. A. K. Duranni, D. C. Hasko, D. A. Williams, J. Vac. Sci. Technol. B 20 (2002) 2806-2809.

[310] J. R. Heath, P. K. Kuekes, G. S. Snider, R. S. Williams, Science 280 (1998) 1716-1721.

[311] M. Kirihara and K. Taniguchi, Jpn. J. Appl. Phys. (pt. 1) 36 (1997) 4172-4175.

[312] C. Gerousis, S. M. Goodnick, W. Porod, Int. J. of Circuit Theory and Appl. 28 (2000) 523-535.

[313] J. C. Da Costa *et al.*, Analog Integrated Circuits and Signal Processing 24 (2000) 59-71.

[314] H. Park *et al.*, Nature 407 (2000) 57-60.

[315] J. Chen *et al.*, in: H. Morkoç, ed. Advanced Semiconductor and Organic Nano-Techniques, Part III (Elsevier, Amsterdam, 2003) 43-187.

[316] Ö. Türel, I. Muckra, and K. Likharev, Possible Nanoelectronic Implementation of Neuromorphic Networks, accepted for presentation at the Int. Joint Conference on Neural Networks (Portland, OR, July 2003); preprint available on the Web at `http://rsfq1.physics.sunysb.edu/~likharev/nano/Portland.pdf`.